Cuadernos 39

La naturaleza de la moralidad. Una introducción a la ética

Gilbert Harman

Universidad Nacional Autónoma de México
Instituto de Investigaciones Filosóficas
Programa de Maestría y Doctorado en Filosofía

LA NATURALEZA DE LA MORALIDAD

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

Colección: CUADERNOS

Cuaderno 39

GILBERT HARMAN

LA NATURALEZA DE LA MORALIDAD

Una introducción a la ética

TRADUCCIÓN:
CECILIA HIDALGO
REVISIÓN DE LA TRADUCCIÓN:
EDUARDO RABOSSI

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

UNAM
POSGRADO
Filosofía

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE INVESTIGACIONES FILOSÓFICAS
PROGRAMA DE MAESTRÍA Y DOCTORADO EN FILOSOFÍA
MÉXICO 2009

BJ1012
H3718
2009

Harman, Gilbert
La naturaleza de la moralidad: una introducción a la ética / Gilbert Harman ; traducción: Cecilia Hidalgo ; revisión de la traducción: Eduardo Rabossi. — 2a. ed. rev., 1a. reimpr. — México : UNAM, Instituto de Investigaciones Filosóficas : Programa de Maestría y Doctorado en Filosofía, 2009.
182 p. — (Colección cuadernos / Universidad Nacional Autónoma de México. Instituto de Investigaciones Filosóficas ; cuaderno 39)
Traducción de: The nature of morality. An introduction to ethics
ISBN: 978–968–36–5062–7

1. Ética. I. Hidalgo, Cecilia, tr. II. Rabossi, Eduardo, tr. III. t. IV. Ser.

Título original:
The Nature of Morality. An Introduction to Ethics
© Oxford University Press, Inc.
Londres, 1977

Cuidado de la edición, composición y formación tipográfica:
Laura E. Manríquez

Primera edición en inglés: 1977
Primera edición en español: 1983
D.R. © 1983 Universidad Nacional Autónoma de México
Segunda edición en español (revisada): 1996
D.R. © 1996 Universidad Nacional Autónoma de México
Primera reimpresión: 30 de noviembre de 2009
D.R. © 2009 Universidad Nacional Autónoma de México

Prohibida la reproducción total o parcial por cualquier medio
sin la autorización escrita del titular de los derechos patrimoniales.

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS
Ciudad Universitaria, Delegación Coyoacán,
C.P. 04510, México, Distrito Federal
Tels.: 5622 7437 y 5622 7504; fax: 5665 4991
Correo electrónico: libros@filosoficas.unam.mx
Página web: http://www.filosoficas.unam.mx

PROGRAMA DE MAESTRÍA Y DOCTORADO EN FILOSOFÍA
Ciudad Universitaria, Delegación Coyoacán
C.P. 04510, México, Distrito Federal
Tels.: 5622 1820 Todos los derechos reservados

Impreso y hecho en México
ISBN 978–968–36–5062–7

Para Elizabeth

PREFACIO

Este libro constituye una introducción filosófica a la ética. Difiere de los textos existentes pues enfoca un problema básico acerca de la moralidad: su aparente inmunidad a la contrastación empírica. Otros textos o bien pasan completamente por alto esta dificultad, a fin de concentrarse en discusiones de problemas interesantes pero en gran medida no filosóficos, o tratan la dificultad como una más de las diversas cuestiones altamente técnicas que se plantean en algo denominado "metaética".

La metaética se transformó en una materia durante el periodo de la filosofía lingüística, digamos de 1930 a 1960, cuando muchos filósofos de habla inglesa jugaron con la idea de que la filosofía no podía ser otra cosa que el análisis del lenguaje. Según este punto de vista, la ética filosófica era el análisis del lenguaje de la moral. El tema filosófico de la metaética debía distinguirse del tema no filosófico de la ética normativa. Una teoría ética normativa sería una teoría acerca de lo Correcto y lo Incorrecto, lo Bueno y lo Malo, lo Debido y lo Indebido. Sería una materia no lingüística y, en consecuencia, no filosófica. Por otra parte, una teoría metaética sería una explicación adecuadamente filosófica del significado y la justificación de los juicios morales, como los que pueden hacerse en el curso de la exposición de una teoría ética normativa.

Muchos filósofos piensan actualmente que este tipo de distinción descansa en suposiciones altamente controvertibles y aun posiblemente incoherentes acerca del significado y la justificación. Sin embargo, en ese entonces estas suposiciones fueron tratadas como principios autoevidentes que debían ser dominados por todo estudiante de la materia, y los problemas filosóficos básicos de la ética fueron reemplazados por series de preguntas acerca del significado. Como los

temas más importantes fueron olvidados, la ética se tornó falta de interés y, por el año de 1960, el trabajo en metaética casi había llegado a su fin.

Mientras tanto, las suposiciones de la filosofía lingüística habían sido atacadas. Los filósofos comenzaron a creer nuevamente que la filosofía no precisa restringirse al análisis del lenguaje. Esto significaba que la ética filosófica no tenía que restringirse a la metaética. De hecho, los argumentos de W.V.O. Quine y otros filósofos socavaron efectivamente la supuesta distinción entre la metaética y la ética normativa mostrando que no puede haber una separación real entre cuestiones de sustancia y cuestiones de significado.

No obstante, este último punto no fue apreciado inmediatamente. Lo que sí quedó claro de inmediato fue que la ética filosófica no tenía que ser identificada con la metaética. Menos claro fue, inicialmente, que había que abandonar por completo la distinción entre metaética y ética normativa. En este estadio, los filósofos podrían haber tratado de abordar el problema básico de la moral sin el bagaje metaético que había obstaculizado previamente dicho intento. Sin embargo, al no ver cómo hacerlo, retuvieron la distinción entre la ética normativa y la metaética, y prestaron atención a unas cuantas preguntas de ética normativa, discutiendo al principio las pretensiones del utilitarismo (véase el capítulo 13), después varios principios no utilitaristas, y, más recientemente, los detalles de temas morales particulares tales como la guerra, el aborto y la igualdad. A menudo esto ha sido interesante pero con menos frecuencia ha tenido algo que ver con la filosofía. Y, en los últimos años, los filósofos han tenido poco que decir acerca del problema básico concerniente a la moralidad, la aparente irrelevancia de las pruebas basadas en la observación.

Uno de los resultados ha sido que los cursos introductorios de ética, así como los textos y antologías diseñados para tales cursos, se han vuelto más interesantes pero menos filosóficos. Los estudiantes emplean menos tiempo en pensar en problemas técnicos de metaética y más tiempo en abordar los temas morales del día. Los cursos de ética son más "relevantes" y los estudiantes están más "comprometidos". Incluso puede ser cierto que estos cursos estén haciendo algo bueno; quizá los estudiantes aprenden algo; tal vez sus poderes de análisis mejoran. (¡Probablemente no!) Pero los estudiantes de tales cursos no están efectivamente estudiando filosofía. No se consagran a los temas

filosóficos básicos. En consecuencia, estos cursos y estos libros no son satisfactorios aunque sean "interesantes" y "relevantes".

¿Por qué ha ocurrido esto? En parte, porque la metaética se había tornado muy falta de interés y porque los filósofos de la ética habían tardado en apreciar la verdadera significación de los argumentos que socavaron la antigua filosofía lingüística. Otro factor ha sido el creciente interés de profesores y estudiantes por los temas sociales: quisieron emplear más de su tiempo en la discusión de tales temas y estaban encantados al pensar que podían hacerlo legítimamente en el aula. El efecto desafortunado ha consistido en una declinación del contenido filosófico de los cursos de ética y de los libros diseñados para ser usados en tales cursos.

Espero que este libro ayude a revertir esta tendencia. Me he centrado en una dificultad *filosófica* básica. En los casos en que han surgido otras varias cuestiones, he indicado cómo se relacionan con esta dificultad básica. He evitado los problemas puramente técnicos que no conciernen directamente a la dificultad principal; es demasiado fácil desviarse. En particular, evito simplemente el aparato metaético de una época anterior.

Algunos libros de texto pretenden exponer diversas posiciones filosóficas de una manera neutral, dejando que el lector decida entre ellas. En mi opinión, esto ofrece una impresión engañosa de la filosofía. La filosofía apunta a la comprensión. En ética, por ejemplo, se debe tratar de desarrollar una explicación comprehensiva de la naturaleza de la moral, y esto es lo que he intentado en este libro. Aunque discuto una variedad de concepciones, estará claro cuál es mi propia posición. El libro en su conjunto representa una sostenida defensa de tal posición. Supongo que el lector no resultará convencido; pero espero que se vea llevado a tratar de desarrollar una explicación más adecuada.

Estoy en deuda con Judy Jarvis Thomson, por sus múltiples comentarios y sugerencias acerca de los capítulos 1 al 10; asimismo, agradezco a Jim Anderson, de Oxford University Press, quien me brindó muchos consejos útiles. También estoy en deuda, aunque de una manera diferente, con la beca del American Council of Learned Societies, la cual me permitió disponer del tiempo necesario para escribir este libro.

Para esta obra, en algunas ocasiones he extraído argumentos y ejemplos de "Moral Relativism Defended", *Philosophical Review*, vo-

lumen 84 (1975), páginas 3–22. El capítulo 10 apareció previamente en *Crítica*, volumen VII, número 21 (1975). El capítulo 2 fue publicado en *Diánoia*, número XXI (1957), en la traducción al castellano de Hugo Margáin.

G.H.

Princeton, Nueva Jersey
Septiembre de 1976

1. EL PROBLEMA CON LA ÉTICA

I

ÉTICA Y OBSERVACIÓN

1. *La dificultad básica*

¿Pueden los principios morales ser puestos a prueba y ser confirmados a la manera de los principios científicos? Consideremos el principio según el cual si debes elegir entre la vida de cinco personas y la muerte de una, o la muerte de cinco personas y la vida de una, siempre deberías elegir la vida de cinco personas y la muerte de una en lugar de la otra opción. Fácilmente podemos imaginar ejemplos que parecen confirmar este principio. He aquí uno:

> Eres médico en la sala de emergencia de un hospital a la que fueron llevadas seis víctimas de un accidente. Las seis están en peligro de muerte pero una está peor que las otras. Puedes llegar a salvar a esa persona si dedicas todos tus recursos a ella y dejas morir a las otras. De manera alternativa, puedes salvar a las otras cinco si estás dispuesto a no atender a la persona más seriamente lesionada.

Parecería que en ese caso, tú, el médico, harías lo correcto si salvaras a las cinco y dejaras morir a la otra persona. Por lo tanto, este ejemplo tomado en sí mismo confirma el principio que hemos señalado. Consideremos ahora el siguiente caso.

> En el hospital tienes cinco pacientes que están muriendo, cada uno de los cuales necesita un órgano diferente. Uno precisa un riñón, otro un pulmón, un tercero un corazón, y así sucesivamente. Puedes salvar a los cinco si tomas a una sola persona sana y le sacas el corazón, los pulmones, los riñones, etc., para distribuirlos entre esos cinco pacientes. Tal persona sana está en la habitación 306. Está en el hospital por unos exámenes de rutina. Como has visto los resultados de los exámenes sabes que está

> perfectamente sana y que posee una compatibilidad tisular adecuada. Si no haces nada, tal persona sobrevivirá sin dificultad; sin embargo, los otros pacientes morirán. Los otros cinco pacientes pueden ser salvados sólo si la persona del cuarto 306 es despedazada y se distribuyen sus órganos. En ese caso, moriría uno y se salvarían cinco.

El principio en cuestión nos dice que deberías despedazar al paciente de la habitación 306. Pero en este caso, ciertamente, no tienes que sacrificar a este espectador inocente, ni siquiera para salvar a los otros cinco pacientes. En este caso hemos puesto a prueba y hemos rebatido un principio moral de una manera que puede parecer sorprendente.

Por supuesto, esto fue un "experimento mental". En realidad no comparamos una hipótesis con el mundo. Comparamos un principio explícito con nuestros sentimientos acerca de ciertos ejemplos imaginados. De la misma manera, un físico realiza experimentos mentales a modo de comparar hipótesis explícitas con su "sensación" [*sense*] de lo que ocurriría en ciertas situaciones, una "sensación" que ha adquirido como resultado de su prolongada familiaridad con la teoría vigente. Pero las hipótesis científicas también pueden ponerse a prueba en experimentos reales, en el mundo.

¿Pueden los principios morales ser puestos a prueba de la misma manera, en el mundo? Puedes observar a alguien haciendo algo, pero ¿puedes percibir la corrección o incorrección de lo que hace? Si doblas una esquina y ves a un grupo de jovenzuelos pillos verter gasolina sobre un gato y prenderle fuego, no necesitas *concluir* que lo que están haciendo es incorrecto, no tienes nada que resolver; puedes *ver* que es incorrecto. Pero, ¿se debe tu reacción a la incorrección efectiva de lo que ves o es simplemente una reflexión de tu "sentido" moral, un "sentido" que has adquirido quizá como resultado de tu educación moral?

2. *La observación*

El tema es complicado. No hay observaciones puras. Las observaciones están siempre "cargadas de teoría". Lo que percibes depende en alguna medida de la teoría que sostienes, consciente o inconscientemente. Ves a algunos niños verter gasolina sobre un gato y prenderle fuego. Para ver realmente eso, debes poseer mucho conocimiento, conocer un considerable número de objetos, saber cómo es la gente: que la gente a lo largo de su vida atraviesa los estadios de recién nacido, bebé, niño, adolescente, adulto. Tienes que saber que los animales y

en particular los gatos, son de carne y hueso. Debes tener alguna idea acerca de la vida. Debes saber qué es la gasolina, qué es incendiarse, y mucho más. En un sentido, lo que "ves" es una configuración luminosa sobre tu retina, una distribución cambiante de manchas, aunque incluso eso es teoría, y nunca podrías describir adecuadamente lo que ves en ese sentido. En otro sentido, ves lo que ves merced a las teorías que sostienes. Si esas teorías cambian verás otra cosa, dada la misma configuración luminosa.

De modo semejante, si sostienes un punto de vista moral, ya sea de manera consciente o inconsciente, serás capaz de percibir la corrección o incorrección, la bondad o maldad, la justicia o injusticia. En este aspecto no hay diferencia entre las proposiciones morales y otras proposiciones teóricas. Si hay alguna diferencia, debe de encontrarse en otra parte.

La observación depende de la teoría porque la percepción involucra la formación de una creencia como un resultado bastante directo de observar algo; puedes formar una creencia sólo si comprendes los conceptos relevantes y un concepto es lo que es en virtud del papel que tiene en alguna teoría o sistema de creencias. Reconocer que un niño es un niño es emplear, consciente o inconscientemente, un concepto que está definido por su ubicación en el marco de los estadios de la vida humana. De manera similar, incendiarse es un concepto vacío al margen de sus conexiones teóricas con los conceptos de calor, destrucción, humo y fuego.

Los conceptos morales —Correcto e Incorrecto, Bueno y Malo, Justo e Injusto— también tienen un lugar en tu teoría o sistema de creencias y son los conceptos que son, en virtud de su contexto. Si decimos que la observación acaece siempre que una opinión es el resultado directo de la percepción, tenemos que admitir que hay observación moral, pues tal opinión puede ser tan fácilmente una opinión moral como de cualquier otro tipo. En este sentido, la observación puede usarse para confirmar o rebatir teorías morales. Las opiniones observacionales con las que, en este sentido, te encuentras pueden estar en concordancia o en conflicto con tus principios morales conscientemente explícitos. Cuando están en conflicto, debes elegir entre tu teoría explícita y la observación. En ética, como en la ciencia, algunas veces hay que optar por la teoría y decir que cometiste un error en la observación o que estabas prejuiciado, o lo que sea; y algunas otras tienes que optar por la observación y modificar tu teoría.

En otras palabras, tanto en ciencia como en ética se invocan principios generales para explicar casos particulares y, en consecuencia, tanto en ciencia como en ética, los principios generales que aceptamos pueden contrastarse apelando a juicios particulares según los cuales ciertas cosas son correctas o incorrectas, justas o injustas, etc.; y estos juicios son análogos a los juicios perceptuales directos acerca de hechos.

3. *Pruebas basadas en la observación*

Sin embargo, la observación desempeña en la ciencia un papel que no parece tener en la ética. La diferencia es que necesitas efectuar suposiciones acerca de ciertos hechos físicos para explicar la ocurrencia de las observaciones que dan apoyo a una teoría científica, pero no parece necesario efectuar suposiciones acerca de algún hecho moral para explicar la ocurrencia de las así llamadas observaciones morales de las que he estado hablando. En el caso moral, parecería que sólo necesitas realizar suposiciones acerca de la psicología o de la sensibilidad moral de la persona que realiza la observación moral. En el caso científico, la teoría se contrasta con el mundo.

El punto es sutil pero importante. Consideremos a un físico que está realizando una observación para poner a prueba una teoría científica. Al ver una huella de vapor en una cámara de niebla, el físico piensa "Ahí va un protón". Supongamos que ésta es una observación en el sentido relevante, es decir, un juicio inmediato formulado como respuesta a la situación sin que haya tenido lugar ningún razonamiento consciente. Supongamos también que la observación confirma su teoría, una teoría que lo ayuda a otorgar significado al propio término "protón" tal como ocurre con su juicio observacional. Tal confirmación descansa en inferir una explicación. El físico puede tomar la realización de la observación como una prueba confirmatoria de su teoría sólo en la medida en que es razonable explicar su realización de la observación mediante el supuesto de que no sólo él está en cierto estado psicológico, dada la teoría que acepta y sus creencias acerca del aparato experimental, sino además que efectivamente hubo un protón que atravesó la cámara de niebla causando la huella de vapor que él vio como un protón. (Ésta es una prueba a favor de la teoría, en la medida en que ésta pueda explicar mejor que otras teorías alternativas que el protón está allí.) Pero si la realización de la observación pudiera igualmente haber sido explicada sólo por su estado psicológi-

co sin necesitar ninguna suposición acerca de un protón, entonces la observación no habría sido una prueba de la existencia de tal protón y, en consecuencia, no habría sido una prueba a favor de la teoría. La realización de la observación da apoyo a la teoría solamente porque, para explicar la realización de la observación, es razonable suponer algo acerca del mundo más allá de las suposiciones hechas acerca de la psicología del observador. En particular, es razonable suponer que había un protón atravesando la cámara de niebla, lo cual causó la huella de vapor.

Comparemos este caso con uno en el que formulas un juicio moral de manera inmediata y sin razonamiento consciente, digamos, que los niños hacen mal en atacar con fuego al gato o que el doctor haría mal si despedazara al paciente sano para salvar a los cinco pacientes moribundos. Para explicar la formulación del primero de estos juicios sería razonable suponer, quizá, que los niños efectivamente están virtiendo gasolina sobre el gato y que los estás viendo hacerlo. Pero en ningún caso existe una razón obvia para suponer nada acerca de "hechos morales" tales como que es realmente incorrecto atacar al gato con fuego o despedazar al paciente de la habitación 306. Efectivamente, una suposición acerca de hechos morales parecería ser completamente irrelevante para explicar la formulación que realizas del juicio. Parecería que todo lo que necesitamos suponer es que tienes ciertos principios morales más o menos bien articulados que se reflejan en los juicios que realizas, basado en tu sensibilidad moral. Parece ser completamente irrelevante para nuestra explicación que tu juicio intuitivo inmediato sea verdadero o falso.

La observación de un suceso puede proporcionar pruebas observacionales a favor o en contra de una teoría científica en el sentido de que la verdad de esa observación puede ser relevante para una explicación razonable de por qué se realizó esa observación. Una observación moral no parece ser, en el mismo sentido, una prueba observacional a favor o en contra de cualquier teoría moral, ya que la verdad o falsedad de la observación moral parece ser completamente irrelevante para cualquier explicación razonable de por qué se realizó esa observación. El hecho de que se realice la observación de un evento en el momento en que se realiza no sólo es una prueba acerca del observador sino también acerca de los hechos físicos. El hecho de que realizaras una observación moral particular cuando lo hiciste no parece ser una prueba de hechos morales, sólo es una prueba acerca de ti y acerca de tu sensibilidad moral. Los hechos acerca de los proto-

nes pueden afectar lo que observas, ya que un protón atravesando una cámara de niebla puede causar una huella de vapor que refleja luz en tu ojo de modo que, dado tu entrenamiento científico y tu tendencia psicológica, te conduce a juzgar que lo que ves es un protón. Pero no parece haber ninguna manera según la cual la efectiva corrección o incorrección de una situación dada pueda tener algún efecto sobre tu aparato perceptivo. En este aspecto, la ética parece diferir de la ciencia. Al considerar si los principios morales pueden ayudar a explicar las observaciones es importante, en consecuencia, advertir una ambigüedad en la palabra "observación". Ves a los niños atacar al gato con fuego e inmediatamente piensas "Eso es incorrecto". En un sentido, tu observación consiste en que lo que los niños están haciendo es incorrecto. En otro sentido, tu observación consiste en ese pensamiento. Los principios morales podrían explicar las observaciones en el primer sentido pero no en el segundo. Ciertos principios morales podrían ayudar a explicar por qué fue *incorrecto* por parte de los niños atacar al gato con fuego, pero los principios morales no parecen ser de ayuda en la explicación de *tu pensar* que fue incorrecto. En el primer sentido de "observación" los principios morales pueden ser puestos a prueba mediante la observación —"Que este acto sea incorrecto es una prueba de que causar sufrimiento innecesario es incorrecto". Pero en el segundo sentido de "observación" los principios morales no pueden ser puestos a prueba de manera clara mediante la observación, dado que no parecen ayudar a explicar las observaciones en el segundo sentido de "observación". Los principios morales no parecen ayudar a explicar que observes lo que observas.

Por supuesto, si ya posees el principio moral según el cual es incorrecto causar sufrimiento innecesario, puedes tomar el ver a los niños atacando al gato con fuego como una prueba observacional de que están haciendo algo incorrecto. De manera similar, puedes suponer que ver la huella de vapor es una prueba observacional de que un protón está atravesando la cámara de niebla, si ya posees la teoría física relevante. Pero existe una importante diferencia manifiesta entre ambos casos. En el caso científico, realizar la observación es prueba en sí misma para la teoría física porque la teoría física explica el protón, que explica la huella, que explica tu observación. En el caso moral, realizar la observación no parece ser una prueba a favor del principio moral relevante porque ese principio no parece explicar la observación. En la moralidad, la cadena explicativa que va del principio a la observación parece estar rota. El principio moral puede "explicar"

por qué es incorrecto por parte de los niños atacar con fuego al gato. Pero la incorrección de tal acto no parece explicar el acto en sí, que tú observas. La cadena explicativa parece romperse de tal manera que ni el principio moral ni la incorrección del acto pueden ayudar a explicar por qué observas lo que observas.

En este caso parece ser necesaria una restricción. Quizá los niños atacan perversamente con fuego al gato simplemente "porque es incorrecto". Aquí, en principio, puede parecer que la incorrección efectiva del acto sí nos ayuda a explicar por qué lo hacen y, en consecuencia, también ayuda a explicar indirectamente por qué observas lo que observas, del mismo modo en que la teoría física, al explicar por qué el protón está produciendo la huella de vapor, ayuda indirectamente a explicar por qué el observador observa lo que observa. Pero pensándolo bien tenemos que aceptar que esto es probablemente una ilusión. Lo que explica el acto de los niños no es claramente la incorrección efectiva del mismo sino, más bien, su creencia de que es incorrecto. La corrección o incorrección efectiva de su acto no parece tener nada que ver con la razón de su proceder.

La prueba observacional desempeña en la ciencia un papel que no parece tener en la ética, pues los principios científicos pueden ser justificados en último extremo por su papel en la explicación de las observaciones, en el segundo sentido de observación, esto es, mediante su papel explicativo. Aparentemente los principios morales no pueden ser justificados de la misma manera. Parece ser verdad que no puede haber una cadena explicativa entre los principios morales y las observaciones particulares de la manera en que tal cadena puede existir entre los principios científicos y las observaciones particulares. Concebida como una teoría explicativa, la moralidad, a diferencia de la ciencia, parece estar aislada de la observación.

No es que toda hipótesis científica legítima sea susceptible de contrastación observacional directa. Ciertas hipótesis acerca de los "agujeros negros" en el espacio no pueden contrastarse directamente, por ejemplo, porque desde el interior de un agujero negro no se emite ninguna luz. En tal caso, la conexión con la observación es indirecta. Y hay muchos ejemplos semejantes. No obstante, visto a grandes rasgos, entre la ciencia y la ética existe la diferencia manifiesta que hemos advertido. El ámbito científico es accesible a la observación de una manera que no lo es el ámbito moral.

4. *La ética y la matemática*

Quizá la ética no deba ser comparada con la física sino con la matemática. Quizá un principio moral tal como "Debes cumplir tus promesas" es confirmado o rebatido de la manera (sea la que fuere) en que lo es un principio matemático tal como "5 + 7 = 12". La observación no parece desempeñar en matemáticas el papel que desempeña en física. No percibimos ni podemos percibir números, por ejemplo, dado que no podemos estar en contacto causal con ellos. Ni siquiera entendemos en qué consistiría estar en contacto causal con el número 12, digamos. Las relaciones entre los números no pueden tener más efecto sobre nuestro aparato perceptivo que el que pueden tener los hechos morales.

Sin embargo, la observación es relevante para la matemática. Al explicar las observaciones que dan apoyo a una teoría física, los científicos, de manera típica, apelan a los principios matemáticos. Por un lado, uno nunca parece necesitar apelar de esta manera a los principios morales. Dado que una observación es una prueba a favor de lo que mejor la explica, y dado que la matemática a menudo figura en las explicaciones de las observaciones científicas, existe una prueba observacional indirecta a favor de la matemática. No parece haber una prueba observacional, ni siquiera indirecta, a favor de los principios morales básicos. Al explicar por qué se han realizado ciertas observaciones nunca parecemos usar exclusivamente suposiciones morales. En este aspecto, entonces, la ética parece diferir no sólo de la física sino además de la matemática.

En lo que sigue, consideraremos un número de posibles respuestas al hecho manifiesto de que la ética está aislada de la contrastación observacional de una manera en la que la ciencia no lo está. Algunas de estas respuestas pretenden que existe una distinción de este tipo entre la ciencia y la ética y tratan de decir cuáles son sus implicaciones. Otras niegan que exista una distinción de este tipo entre la ciencia y la ética y argumentan que la ética no está efectivamente exenta de contrastación observacional de la manera en que parece estarlo.

NOTA ACERCA DE LECTURAS ADICIONALES

Para un argumento breve que distingue el papel de la prueba observacional en ética y en ciencia, veáse R.M. Hare, *Freedom and Reason*, Oxford University Press, Oxford, 1963, pp. 1–3.

Alan Gewirth advierte algunas complicaciones en "Positive 'Ethics' and Normative 'Science' ", *Philosophical Review*, vol. 69, 1960.

Acerca del carácter "cargado de teoría" de la observación, véase Norwood Russell Hanson, *Patterns of Discovery*, Cambridge University Press, Cambridge, 1958, cap. 1.

El papel de la explicación en la inferencia se discute en Gilbert Harman, "Inference to the Best Explanation", *Philosophical Review*, vol. 74, 1965.

La tesis de que puede haber conocimiento intuitivo de las verdades morales se examina en P.F. Strawson, "Ethical Intuitionism", *Philosophy*, vol. 24, 1949.

Paul Benacerraf discute problemas acerca del conocimiento matemático en "Mathematical Truth", *Journal of Philosophy*, vol. 70, 1973.

II

NIHILISMO Y NATURALISMO

1. *El nihilismo moral*

Hemos visto que la prueba observacional desempeña en la ciencia y en la matemática un papel que no parece tener en la ética. Las hipótesis morales no ayudan a explicar por qué la gente observa lo que observa. Por lo tanto, la ética es problemática y el nihilismo debe ser tomado en cuenta seriamente. El nihilismo es la doctrina según la cual no hay hechos morales, ni verdades morales, ni conocimiento moral. Esta doctrina puede dar cuenta de por qué la referencia a los hechos morales no ayuda a explicar las observaciones, sobre la base de que lo que no existe nada puede explicar.

Una versión extrema del nihilismo sostiene que la moralidad es simplemente una ilusión: jamás algo es correcto o incorrecto, justo o injusto, bueno o malo. En esta versión deberíamos abandonar la moralidad, tal como un ateo abandona la religión luego de decidir que los hechos religiosos no pueden ayudar a explicar las observaciones. Algunos nihilistas extremos han señalado incluso que la moralidad es meramente un remanente supersticioso de la religión.

Tal nihilismo extremo es difícil de aceptar. Implica que las constricciones morales no existen; que todo está permitido. Tal como observa Dostoievsky, implica que nada hay de incorrecto en que asesines a tu padre. También implica que la esclavitud no es injusta y que los campos de exterminio de Hitler no fueron inmorales. Éstas no son conclusiones fáciles de aceptar.

Por supuesto, esto no refuta el nihilismo extremo. El nihilismo no se propone reflejar nuestros puntos de vista ordinarios; y el hecho de que sea difícil de creer no significa que tenga que ser falso. En un momento de la historia del mundo, a la gente le resultaba difícil creer

que la Tierra era redonda; sin embargo, la Tierra era redonda. Una persona sinceramente religiosa no podría llegar a creer con facilidad que Dios no existe: éste no es un argumento contra el ateísmo. El nihilismo extremo es un punto de vista posible y merece ser tomado en cuenta seriamente.

Por otra parte, también vale la pena señalar que el nihilismo extremo no es una consecuencia automática de la aseveración de que aparentemente los hechos morales no pueden ayudarnos a explicar las observaciones. Aunque esto es una base para el nihilismo, hay versiones más moderadas de él. No todas las versiones implican que la moralidad es una ilusión y que los juicios morales deben ser abandonados a la manera en que el ateo abandona los juicios religiosos. De este modo, un nihilismo más moderado sostiene que el propósito de los juicios morales no es describir el mundo sino expresar nuestros sentimientos morales o servir como imperativos que dirigimos a nosotros mismos y a los demás. Desde esta óptica, la moralidad no está socavada por su aparente incapacidad para explicar las observaciones, pues esperar que los juicios morales nos ayuden a explicar las observaciones es confundirse acerca de la función de la moralidad. Es como si se esperara explicar las observaciones exclamando "¡Ay!" u ordenando "¡Cierra la puerta!"

El nihilismo moderado es más fácil de aceptar que el nihilismo extremo. Nos permite conservar la moralidad y continuar efectuando juicios morales. Esto no implica que no haya nada incorrecto en que asesines a tu padre, seas dueño de esclavos o establezcas campos de exterminio. Dado que desaprobamos estas actividades podemos, según el nihilismo moderado, expresar legítimamente nuestra desaprobación diciendo que son incorrectos.

No obstante, el nihilismo moderado también está en conflicto con el sentido común, aunque el conflicto es menos ostensible. Afirmar, como también afirma el nihilista moderado, que no hay hechos morales, ni verdades morales, ni conocimiento moral es afirmar algo que va contra una gran parte de lo que ordinariamente decimos y pensamos. Si alguien señala que es incorrecto por parte de los miembros de la Unión de Contribuyentes de Oregón haber secuestrado a Sally Jones con el objeto de extorsionar a su padre, Austin P. Jones, y tú estas de acuerdo, expresarás tu acuerdo diciendo "¡Es *verdad*!" De manera similar, al decidir qué hacer en una ocasión particular, dices cosas tales como ésta: *Sé* que debería cumplir con lo que le prometí a Herbert, pero realmente hoy desearía ir a la playa". Ordinariamente hablamos

de los juicios morales como verdaderos o falsos; y hablamos como si conociéramos ciertas verdades morales y no otras.

El nihilismo, extremo o moderado, entonces, está en conflicto con las maneras ordinarias de hablar y de pensar. Aunque tal conflicto no refuta una teoría, tenemos que preguntarnos si podemos ajustar la aseveración acerca de la ética y la observación sin tener que abandonar nuestros puntos de vista ordinarios y suscribiendo al mismo tiempo alguna forma de nihilismo.

2. *Reducciones*

Nuestra discusión previa sugiere el siguiente argumento en favor del nihilismo moral:

> Las hipótesis morales nuncan ayudan a explicar por qué observamos algo. Por lo tanto no tenemos pruebas en favor de nuestras opiniones morales.

El argumento depende de esta suposición:

> Podemos tener pruebas para hipótesis de cierto tipo sólo si tales hipótesis ayudan algunas veces a explicar por qué observamos lo que observamos.

Pero esta suposición es demasiado fuerte. Las hipótesis acerca del ciudadano estadounidense medio nunca pueden ayudarnos a explicar por qué observamos algo acerca de un estadounidense particular, pero sí podemos obtener una prueba en favor de tales hipótesis obteniendo pruebas en favor de las hipótesis acerca de los ciudadanos de Estados Unidos. La razón es que los hechos acerca del ciudadano estadounidense medio son definibles en términos de hechos acerca de ciudadanos estadounidenses. Los hechos del primer tipo están construidos a partir de los hechos del segundo tipo y, en consecuencia, son reducibles a ellos. Aunque las suposiciones acerca de los hechos morales no ayuden directamente a explicar las observaciones, puede ser que los hechos morales puedan ser reducidos a otros tipos de hechos y que las suposiciones acerca de estos hechos ayuden a explicar las observaciones. En ese caso, podría haber pruebas a favor de las suposiciones acerca de los hechos morales.

Tomemos otro ejemplo. Podríamos ser capaces de dar cuenta de la percepción del color sin suponer que los objetos efectivamente tienen colores. Pues podríamos estar en condiciones de explicar cómo

reflejarán luz de una longitud de onda particular los objetos cuyas superficies tienen ciertas características físicas; entonces esa luz golpea la retina del ojo del observador, afectándolo de un modo que puede ser descrito por una psicología neurofisiológica adecuada. Es decir, podríamos estar en condiciones de explicar la percepción del color enteramente en términos de las características físicas de los objetos percibidos y de las propiedades de la luz junto con una explicación del aparato perceptivo del observador. Esto no probaría que no hay hechos acerca de colores; sólo mostraría que los hechos acerca de los colores no son hechos adicionales más allá de los hechos físicos y psicológicos. Si pudiéramos explicar de este modo la percepción del color, podríamos concluir que los hechos acerca del color son de alguna manera reducibles a hechos acerca de características físicas de los objetos percibidos, hechos acerca de la luz y hechos acerca de la psicología y del aparato perceptivo de quienes perciben. Podríamos considerar si los hechos morales, de manera similar, son conformables a partir de ciertos hechos que pueden ayudarnos a explicar las observaciones, o bien son reducibles a ellos.

3. *El naturalismo ético: el funcionalismo*

Ésta es ciertamente una sugerencia verosímil para ciertos hechos evaluativos no morales. Consideremos, por ejemplo, lo que está involucrado en que algo sea una buena cosa de su tipo, un buen cuchillo, un buen reloj o un buen corazón. Ciertas funciones están asociadas a estos tipos de cosas. Un cuchillo es algo que se usa para cortar; un reloj se usa para registrar la hora; un corazón es el órgano que bombea la sangre. Además, algo es una buena cosa del tipo relevante en la medida en que satisfaga adecuadamente su función propia. Un buen cuchillo corta bien; un buen reloj da la hora justa; un buen corazón bombea la sangre con la presión adecuada sin tropiezos. Usemos la letra "*K*" para designar un tipo de cosa. En estos casos, un buen *K* es un *K* que satisface adecuadamente su función. El hecho de que algo sea un buen *K* o de que no lo sea es una cuestión fáctica, pues el hecho de que *K* tenga o no esa función es una cuestión fáctica, así como es una cuestión fáctica el hecho de que una determinada cosa satisfaga adecuadamente o no esa función.

Más aún, un *K* debería de satisfacer su función. Si no lo hace, algo anduvo mal. En consecuencia, es una cuestión fáctica que un determinado *K* de este tipo sea como debe ser y haga lo que debe

hacer y es una cuestión fáctica que algo ande mal con un *K* de este tipo. Un cuchillo debe estar afilado de manera que corte bien. Hay algo de malo con un corazón que no bombea la sangre sin tropiezos.

Hay aquí, por supuesto, dos casos algo diferentes: artefactos, como los relojes y los cuchillos, y partes de sistemas naturales, como los corazones. Las funciones de los artefactos están determinadas por sus fabricantes y sus usuarios. Las funciones de las partes de sistemas naturales están determinadas por su papel en el sostenimiento de dichos sistemas. De tal modo, en cada caso es una cuestión fáctica determinar cuál es la función relevante de un *K*.

Consideremos a continuación un grupo algo diferente de casos: una buena comida, un buen chapuzón, un buen momento. Podemos extender el punto y decir que las comidas, los chapuzones y los momentos tienen funciones o propósitos; pero sería más adecuado decir que pueden responder a ciertos intereses. Juzgamos que las comidas particulares, los chapuzones o los momentos son buenos porque responden a los intereses relevantes. Allí donde son relevantes diferentes clases de intereses encontramos ambigüedad: "una buena comida" puede significar una comida nutritiva o una comida sabrosa.

En este grupo "debería" e "incorrecto" están usados como antes. Una buena comida debe ser balanceada (o sabrosa). Algo anda mal con un filete que no es tierno y jugoso.

Los casos más complejos involucran los papeles que puede tener una persona de una manera u otra: un buen campesino, un buen soldado, un buen maestro, un buen ciudadano, un buen ladrón. Una persona es evaluada en términos de funciones, de los papeles que desempeña y de diversos intereses, de una manera que es difícil especificar. Aquí también las palabras "debería" e "incorrecto" son relevantes como antes. Durante una batalla, decimos que un soldado debe obedecer a sus oficiales superiores sin cuestionarlos. Es incorrecto por parte de un maestro mostrar favoritismos. Un ladrón debe usar guantes.

Algunos tipos de cosas no están asociados con funciones, propósitos o clases de intereses; por ejemplo, las piedras en sí no lo están. En consecuencia, no tiene sentido preguntarnos fuera de un contexto específico si algo es una buena piedra. Sólo podemos contestar tal pregunta en relación con los intereses que podríamos tener en usos posibles de la piedra. Por ejemplo, podría ser una buena piedra para usarla como pisapapeles; pero si es para usarse como tope de una puerta, quizá debería ser más pesada.

Los juicios evaluativos relevantes son fácticos. Los hechos son hechos naturales aunque un tanto complejos. Juzgamos que algo es bueno o malo, que es correcto o incorrecto, que debe o no debe tener ciertas características o hacer ciertas cosas, en relación con un haz de intereses, papeles y funciones. Podemos abreviar esto diciendo que algún *X* es bueno en la medida en que responda adecuadamente a los intereses relevantes. Especificar estos intereses es especificar en qué caso es bueno *X* [*what X is good as*]. De modo semejante, una persona *P* debería hacer *H* si y sólo si el hacer *H* por parte de *P* responde a los intereses relevantes.

Éste es un análisis realista para muchos casos y señala cómo deben construirse los hechos evaluativos a partir de hechos observables aun cuando los hechos evaluativos en sí mismos no figuren en las explicaciones de las observaciones. El hecho de que mi reloj sea bueno puede no explicar nada acerca de mis observaciones de él; pero el hecho de que registre exactamente la hora ayuda a explicar su concordancia continua con los anuncios de la hora en la radio y quizá la bondad de mi reloj consista en hechos de este tipo.

Pero cuando este tipo de análisis se aplica a la ética se manifiesta un problema. Consideremos el caso en el que tú eres el médico que puede salvar a los cinco pacientes despedazando al paciente sano de la habitación 306 y distribuyendo sus órganos entre los otros pacientes, o que puedes no hacer nada y dejar morir a los otros cinco pacientes. El problema consiste en que en cada caso estarías satisfaciendo ciertos intereses y no estarías satisfaciendo otros. Los intereses de los cinco pacientes moribundos están en conflicto con los intereses del paciente sano de la habitación 306. La pregunta moral concierne a lo que deberías hacer, tomando en cuenta todos los intereses. Tal como lo vimos anteriormente, nuestro juicio intuitivo consiste en que no deberías sacrificar al paciente de la habitación 306 para salvar a los otros cinco pacientes. ¿Es éste un juicio fáctico? Si suponemos que es un hecho que no debas sacrificar al paciente de la habitación 306, ¿cómo está relacionado este hecho con los hechos que pueden ayudarnos a explicar las observaciones? No es obvio cómo podemos extender nuestro análisis para abarcar este tipo de casos.

En realidad, el problema no es peculiar de la ética. Un reloj pesado, a prueba de agua, a prueba de golpes, que puede soportar una presión considerable, ¿es mejor o peor reloj que uno más liviano, elegante, delicado, sin aquellos rasgos? ¿Es mejor o peor que otra una maestra que hace infelices a sus alumnos al enseñarles más?

En alguna medida, nuestra dificultad en estos casos reside en la vaguedad de nuestros estándares para los relojes y los maestros. A menudo podemos resolver la vaguedad especificando los intereses relevantes. El reloj pesado es mejor para bucear. El reloj más liviano es mejor para ocasiones sociales, fuera del agua. En el caso de la evaluación de los maestros debemos decidir qué queremos de los maestros —quizá que sus alumnos aprendan una cantidad mínima y, cuando aprendan al menos eso, que no sean desdichados. Pero aunque se ofrezcan, de esta manera, especificaciones ulteriores de nuestros intereses en los relojes y en los maestros, puede no existir un hecho tal como qué reloj o qué maestro sea mejor; no porque éstas no sean preguntas fácticas sino por la vaguedad de los estándares. Las preguntas fácticas son fácticas aun cuando no puedan ser contestadas en virtud de su vaguedad. (¿Está abierta o cerrada una puerta cuando está levemente entreabierta?) Además, aun en los casos en los que sentimos intuitivamente que un reloj o un maestro es mejor, podemos no ser capaces de especificar muy claramente los intereses, funciones y papeles en referencia a los cuales uno es mejor que otro en tanto reloj o maestro. Asimismo, puede muy bien ser un hecho que uno sea mejor —un hecho construido de una manera que sólo podemos especificar vagamente a partir de hechos de un tipo que pueden ayudarnos a explicar las observaciones.

De manera semejante, *puede* ser que los hechos morales, como el hecho de que no se debe sacrificar al paciente sano de la habitación 306 para salvar a los otros cinco pacientes, puedan construirse de un modo u otro a partir de hechos de un tipo que pueda explicar las observaciones, aunque sólo podamos indicar vagamente los papeles, los intereses y las funciones relevantes.

Esto vindicaría el naturalismo ético, que es la doctrina según la cual los hechos morales son hechos de la naturaleza. El naturalismo en tanto visión general es la tesis razonable según la cual *todos* los hechos son hechos de la naturaleza. Por supuesto, podemos aceptar el naturalismo en general sin comprometernos con el naturalismo ético, dado que se puede ser nihilista y negar que haya hechos morales, tal como podemos negar que haya hechos religiosos. Los naturalistas pueden ser nihilistas éticos o naturalistas éticos. La pregunta consiste en cómo decidimos entre el nihilismo ético y el naturalismo ético, y no hay una respuesta simple. Si fuera verosímil un análisis de los hechos morales en tanto hechos acerca de funciones, papeles e intereses, ello sería un argumento poderoso a favor del naturalismo ético.

Pero las funciones, papeles e intereses relevantes a lo sumo pueden ser indicados de manera vaga, por lo tanto, el análisis propuesto es difícil de evaluar. Queda el nihilismo como posibilidad.

4. *El argumento de la pregunta pendiente* [open question]

Por otra parte, los argumentos generales contra el naturalismo ético, y a favor del nihilismo, tampoco son concluyentes. Por ejemplo, los nihilistas moderados arguyen que los naturalistas interpretan equivocadamente la función de los juicios morales, que no consiste en describir los hechos (dicen) sino más bien en expresar la aprobación o desaprobación del hablante. En consecuencia, los nihilistas moderados dicen que el naturalismo ético involucra una "falacia naturalista". Pero, como veremos, la evaluación de esta posición nihilista moderada es también bastante compleja.

Un naturalista ético sostiene que hay hechos morales que pueden ser "reducidos" a hechos naturales de un tipo que podría explicar las observaciones, tal como los hechos acerca del color pueden ser reducidos a hechos acerca de las características físicas de los objetos, de las propiedades de la luz y del aparato perceptivo de un observador. He aludido a un modo en el que un naturalista ético podría intentar tal reducción apelando a funciones, papeles e intereses. También hay otros modos: podría, por ejemplo, tratar de desarrollar una teoría del "observador ideal" de hechos morales en analogía con la sugerida teoría de los hechos de color. Consideraremos tal teoría del observador ideal posteriormente, en el capítulo 4. Y también son posibles otros tipos de naturalismo ético. Ahora bien, algunos nihilistas moderados creen que hay un argumento perfectamente general que puede usarse para mostrar, de una vez y para siempre, que toda versión del naturalismo ético debe fracasar. Éste es el denominado "argumento de la pregunta pendiente". En ética toda reducción naturalista tendría la forma "*P* debería hacer *H* si y sólo si hacer *H* por parte de *P* presenta las características *C* ", donde las características *C* son características naturalistas de un tipo que puede ayudarnos a explicar las observaciones. Dada la reducción naturalista propuesta, los defensores del argumento de la pregunta pendiente sostienen que queda pendiente la siguiente pregunta:

> Estoy de acuerdo en que hacer *H* por parte de *P* sería para *P* hacer algo que es *C*, pero, ¿debería *P* hacer *H*?

Esta pregunta queda pendiente, dicen los nihilistas moderados, porque describir un acto no es lo mismo que suscribirlo. Según el nihilista moderado, describas al acto como lo describas, no lo has suscrito y, en consecuencia, no has dicho aún si debería ser realizado. En consecuencia, la pregunta formulada es (afirman) una pregunta pendiente de una manera en que no lo es la siguiente pregunta:

> Estoy de acuerdo en que *P* debería hacer *H*, pero, ¿debería *P* hacer *H*?

Esta pregunta es obviamente tonta. Dado que algo debería hacerse, no puede quedar pendiente la pregunta relativa a si eso debería hacerse. Y dado que la primera pregunta queda pendiente pero la segunda no, debemos concluir que las características naturales de ser un acto que es *C* no pueden equipararse con las características morales de ser un acto que debería ser realizado.

Una dificultad con este argumento es que tiene que mostrarse que la primera pregunta siempre está pendiente. Si al argumentar contra el naturalista ético, un nihilista ético sólo dice que describir un acto en términos de la posesión de ciertas características no cuenta como suscribir el acto en el sentido de decir que debería ser realizado, está incurriendo en una petición de principio. Por ejemplo, no es obvio que la siguiente pregunta quede pendiente en el sentido relevante:

> Estoy de acuerdo en que, si *P* hace *H*, *P* satisfará los intereses relevantes, pero, ¿debería *P* hacer *H*?

Por supuesto que una parte del problema es que los "intereses relevantes" no están especificados de una manera naturalista precisa. Sin embargo, no es obvio que si así fueran especificados, la pregunta quedaría pendiente.

Más importante, quizá, es el hecho de que, tal como está, el argumento de la pregunta pendiente es inválido. Podría usarse un argumento análogo contra quien ignorara la composición química del agua para "probarle" que el agua no es H_2O. Esta persona estaría de acuerdo en que no es una pregunta pendiente la de si el agua es agua, sino que está pendiente, al menos para él, la pregunta acerca de si el agua es H_2O. Dado que este argumento no mostraría que el agua no es H_2O, el argumento de la pregunta pendiente, tal como está, no puede ser usado en ética para mostrar que ser un acto que debería ser realizado no es lo mismo que tener alguna característica natural *C*.

A menudo el argumento de la pregunta pendiente es presentado como una refutación, no del naturalismo ético en general, sino de una versión más particular a la que podemos denominar naturalismo definicional. El naturalismo definicional supone que los juicios morales son definicionalmente equivalentes a los juicios naturales. Entonces, el argumento de la pregunta pendiente debería mostrar que las definiciones propuestas tienen que ser incorrectas.

Sin embargo, existen varios tipos de definiciones y el argumento de la pregunta pendiente no es relevante respecto de la mayoría de ellas. Por ejemplo, un científico define el agua como H_2O y, tal como hemos visto, el argumento de la pregunta pendiente, aplicado a esta definición, no la refuta.

Presumiblemente, el argumento de la pregunta pendiente está dirigido a quien pretenda que una definición naturalista capta el significado de un término moral en el sentido de que los juicios morales, tal como los usamos ordinariamente, son sinónimos de juicios que describen hechos naturales. Si efectivamente la cuestión de si un acto que es *C* es un acto que debe ser realizado es una pregunta pendiente aun para alguien que conozca el significado de "*C*" y "debe ser realizado", ¿cómo pueden ser sinónimos "*C*" y "debe ser realizado"? Debe mostrarse, no sólo suponerse, no obstante, que la pregunta relevante está siempre pendiente, independientemente de cuáles sean las características naturales *C*.

5. *El naturalismo redefinicional*

Otro tipo de naturalismo definicional en ética no es en realidad una versión del naturalismo ético. Para este punto de vista, nuestra terminología moral es tan vaga, tan poco clara y confusa que haríamos muy bien en reemplazarla por una terminología mejor y más precisa. Por ejemplo, alguien que desarrolla la teoría de que debe hacerse lo que responde a los intereses relevantes podría argumentar que nuestra perspectiva del ejemplo que involucra al paciente de la habitación 306 muestra que nuestras ópticas morales son incoherentes. Podría indicar que deberíamos reemplazar nuestras nociones actuales por conceptos más claros, por ejemplo, definiendo "debería" de modo tal que un acto debería ser realizado si y sólo si maximizara la satisfacción de intereses. Mediante este criterio utilitarista, deberías despedazar al paciente de la habitación 306 para salvar a los otros pacientes. Si es cierto que la definición propuesta no capta el significado ordinario de "debería",

porque juzgamos intuitivamente que se debería proteger al paciente sano de la habitación 306, no estamos juzgando, ciertamente, que ello maximizaría la satisfacción de intereses; en verdad, vemos que no lo haría. Pero una definicion no necesita captar lo que queremos decir ordinariamente. Podemos definir nuestros términos como queramos, siempre que estemos dispuestos a usar esos términos en concordancia con nuestras definiciones. La definición sugerida es relativamente clara y precisa. ¿Qué es una definición mejor?

Esta línea argumentativa es inteligible y no es absurda, aunque no deja de tener sus propias dificultades. Tiene que mostrarse y no sólo suponerse que las nociones morales ordinarias son confusas. Ésta es una alegación discutible. El hecho de que no haya una manera obvia de definir la terminología moral ordinaria de modo preciso no muestra que algo no marche bien con esa terminología. No todo término puede ser definido; cabría la posibilidad de que la terminología moral no pudiera ser reducida a una terminología más simple.

Además, en esta línea argumentativa existe el riesgo de que alguien que la adopta pueda hacer trampa, usando algunas veces "debería" tal como lo ha definido y a veces en su sentido ordinario. El mejor modo de evitar este problema sería prescindir, en su conjunto, de la terminología moral en favor de la terminología utilitarista y, en vez de hablar acerca de lo que la gente debería hacer, hablar en cambio acerca de lo que satisfaría la mayor cantidad de intereses. Pero esto sería renunciar a toda pretensión de naturalismo ético y revelar que se ha adoptado el nihilismo extremo. Ello involucraría negar la existencia de los hechos morales en el sentido ordinario de "moralidad", tal como un naturalista general abandona la religión en el sentido ordinario de "religión".

6. *Por qué es problemática la ética*

Aunque no estamos en condiciones de suponer que el nihilismo, extremo o moderado, es correcto, ahora estamos en condiciones de ver claramente el modo en que la ética es problemática. Nuestro punto inicial en este capítulo fue que los juicios morales no parecen ayudar a explicar las observaciones. Esto nos llevó a preguntarnos si existen los hechos morales, las verdades morales y el conocimiento moral. Vimos que podría haber hechos morales si estos actos fueran reducibles de una manera u otra a otros hechos de un tipo que pudiera ayudar a explicar las observaciones. Pues hemos advertido que hay hechos

acerca del ciudadano estadounidense medio, aunque tales hechos no expliquen en sí mismos las observaciones, ya que tales hechos son reducibles a hechos acerca de ciudadanos estadounidenses que pueden ayudar a explicar las observaciones. De modo similar, hemos advertido que no podríamos decidir que no hay hechos acerca de colores aunque fuéramos capaces de explicar la percepción del color sin apelar a hechos acerca de colores; en cambio supondríamos que los hechos acerca de colores son reducibles a hechos acerca de las superficies físicas de los objetos, de las propiedades de la luz y de la psicología neurofisiológica de los observadores. Por lo tanto, concluimos que no teníamos que aceptar el nihilismo ético simplemente porque los hechos morales no parecen ayudar a explicar las observaciones; en cambio podríamos confiar en una reducción naturalista de los hechos morales.

Teniendo esto presente, consideremos la posibilidad de que los hechos morales puedan ser reducidos a hechos acerca de intereses, papeles y funciones. Concluimos que si lo fueran, la reducción tendría que ser compleja, vaga y difícil de especificar. La ética sigue siendo problemática.

Es cierto que la reducción de los hechos acerca de colores también es compleja, vaga y difícil (probablemente imposible) de especificar. Pero hay una diferencia importante entre los hechos acerca de colores y los hechos morales. Aunque lleguemos a estar en condiciones de explicar la percepción del color apelando a las características físicas de las superficies, las propiedades de la luz y la psicología neurofisiológica de los observadores, nos referiremos *algunas* veces a los colores reales de los objetos al explicar la percepción del color, aunque sea por razones de sencillez. Por ejemplo, explicaremos que algo parece verde porque es amarillo y la luz es azul. Es posible que la referencia al color real del objeto en una explicación de este tipo pueda ser reemplazada por nuestro hablar acerca de las características físicas de la superficie. Pero esto complicaría mucho lo que es una explicación simple y fácilmente comprendida. Es por esto que, aun luego de llegar a ser capaces de ofrecer explicaciones sin referirnos a los colores reales de los objetos, todavía suponemos que los objetos tienen colores reales y que, en consecuencia, los hechos acerca de los colores reales de los objetos son de alguna manera reducibles a hechos acerca de las características físicas de las superficies, etc., aun cuando (probablemente) no seamos capaces de especificar la reducción más que de un modo vago. Continuaremos creyendo que los objetos tienen colores,

pues continuaremos refiriéndonos a los colores reales de los objetos en las explicaciones que daremos en la práctica. Una observación similar no parece valer para los hechos morales. No parece tener sentido, ni siquiera en la práctica, explicar las observaciones morales de alguien mediante la apelación a lo que es realmente correcto o incorrecto, justo o injusto, bueno o malo. Siempre parece ser más exacto explicar las observaciones morales citando hechos acerca de puntos de vista morales y de sensibilidad moral. Por lo tanto, las razones para suponer que hay hechos acerca de colores no corresponden a las razones para pensar que hay hechos morales.

Es cierto que los hechos acerca del ciudadano estadounidense medio no parecen ayudar a explicar las observaciones, ni siquiera en la práctica. En esto, tales hechos son semejantes a los hechos morales. Pero hay esta diferencia. Podemos ofrecer una reducción *precisa* de los hechos acerca del ciudadano estadounidense medio porque podemos definir explícitamente estos hechos en términos de hechos que son de un tipo que puede ayudar a explicar las observaciones. El problema con los hechos morales alegados es que, hasta donde podemos verlo por ahora, no hay una manera simple y precisa de definirlos en términos de hechos naturales.

Estamos dispuestos a suponer que hay hechos acerca de los colores a pesar de no saber precisamente cómo reducirlos, pues en la práctica suponemos que hay hechos tales en muchas de nuestras explicaciones de la percepción del color, aunque en teoría esta suposición es dispensable. Estamos dispuestos a suponer que hay hechos acerca del ciudadano estadounidense medio a pesar de que nunca usamos tal suposición para explicar las observaciones, porque podemos reducir precisamente esos hechos a hechos de un tipo que puede ayudar a explicar las observaciones. Dado que los hechos morales no parecen ser precisamente reducibles ni útiles ni siquiera en la práctica en nuestras explicaciones de las observaciones, sigue siendo problemático que tengamos alguna razón para suponer que hay hechos morales.

NOTA ACERCA DE LECTURAS ADICIONALES

Fyodor Dostoievsky examina las implicaciones del nihilismo moral en sus novelas *Crimen y Castigo* y *Los hermanos Karamazov*.

Friedrich Nietzsche arguye en *El crepúsculo de los ídolos* que no hay hechos morales y que la moralidad tal como ha existido hasta ahora debe ser rechazada (Alianza Editorial, Madrid, 1973).

A.J. Ayer arguye en *Language, Truth and Logic* (Dover, Nueva York, 1950) que los enunciados morales no tienen "significado cognoscitivo".

Para una discusión de la idea de que la palabra "bueno" debe ser caracterizada en términos funcionales, véase Philippa Foot, "Goodness and Choice", *Proceedings of the Aristotelian Society*, volumen suplementario 25 (1961), y Jerrold J. Katz, "Semantic Theory and the Meaning of 'Good' ", *Journal of Philosophy*, vol. 61, 1964.

Acerca de la "falacia naturalista" véase G.E. Moore, *Principia Ethica*, Cambridge University Press, Cambridge, 1903, pp. 6–21 [existe la versión en castellano *Principia Ethica*, trad. Adolfo García Díaz, Instituto de Investigaciones Filosóficas, UNAM]; y William K. Frankena, "The Naturalistic Fallacy", *Mind*, vol. 48, 1939.

2. EL EMOTIVISMO

III

EL EMOTIVISMO COMO UN NIHILISMO MODERADO

1. *El emotivismo: la idea básica*

El nihilismo es la doctrina según la cual no hay hechos morales, ni verdades morales, ni conocimiento moral. El nihilismo moderado dice que el nihilismo no constituye una razón para abandonar la moralidad, dado que la moralidad no describe hechos sino que hace algo más. A menudo se señala que los juicios morales expresan los sentimientos o actitudes de la gente que efectúa tales juicios.

Cuando examinas un desacuerdo moral, puedes ver lo que sugiere que los juicios morales son meramente expresiones de actitudes. Si digo que la Unión de Contribuyentes de Oregon hizo lo correcto al secuestrar a Sally Jones y tú dices que eso fue incorrecto, tú y yo estamos en desacuerdo. Entre otras cosas, estamos en desacuerdo en nuestra actitud hacia el secuestro: yo estoy a favor y tú estás en contra; yo tengo una actitud pro hacia el secuestro y tú tienes una actitud contra. Podemos o no diferir en nuestras creencias acerca de la Unión de Contribuyentes de Oregon, de lo que hicieron, de quién es Sally Jones, de quién es su padre, Austin P. Jones, etc. Pero sea que difiramos o no respecto de lo que creemos, además estamos en desacuerdo en cuanto a nuestras actitudes, y nuestro desacuerdo moral, concerniente a si fue correcto o incorrecto por parte de la Unión de Contribuyentes de Oregon secuestrar a Sally Jones, es este desacuerdo en actitudes. Así, independientemente de que lleguemos a un acuerdo en cuanto a lo que creemos acerca de los hechos, mientras estemos en desacuerdo en nuestras actitudes, nuestro desacuerdo moral persistirá. Y tan pronto como estemos de acuerdo en nuestras actitudes, nuestro desacuerdo moral habrá finalizado, sin importar que persista una diferencia en nuestras creencias.

Podrías decir entonces que "*X* es incorrecto" significa "Desapruebo *X*". Pero ésa no sería una formulación correcta. Mi juicio moral no es *acerca* de mi aprobación o desaprobación; más bien *expresa* mi aprobación o desaprobación. Mi juicio moral es a mi aprobación moral lo que mi juicio fáctico es a mi creencia. Si digo que Sally Jones era una estudiante de segundo año de universidad en el momento en que fue secuestrada y tú dices que estaba en tercer año, nuestro desacuerdo es un desacuerdo en creencia; pero no estamos hablando *acerca* de nuestras creencias sino que estamos expresándolas. La observación "Sally Jones era una estudiante de segundo año" no es equivalente a la observación "Una de mis creencias consiste en que Sally Jones era una estudiante de segundo año". Sin embargo, la observación "Sally Jones era una estudiante de segundo año" expresa la creencia en que Sally Jones era una estudiante de segundo año. De manera similar, si afirmo que la Unión de Contribuyentes de Oregon hizo lo correcto al secuestrar a Sally Jones, lo que digo no es equivalente a *decir que* apruebo que la Unión lo haya hecho, sino que lo que digo expresa mi aprobación del secuestro.

El punto de vista que hemos estado discutiendo se conoce también como "emotivismo": los juicios morales expresan las emociones, sentimientos, actitudes, intenciones o, más generalmente, normas y valores del hablante. Valorar algo no sólo es tener una creencia acerca de algo; es tener una actitud hacia algo. Es estar en favor de algo. Valorar algo es estar en un estado emocional, no en un estado cognoscitivo.

En principio, el emotivismo puede parecer una teoría controvertible y radical. Pero una vez que ha sido completamente elaborada y cuando las objeciones han sido dejadas a un lado, veremos que se convierte en un punto de vista más moderado de lo que parecía al comienzo.

Adviértase, entre otras cosas, que el emotivismo es una versión de lo que antes denominamos nihilismo moderado, el cual sostiene que no hay hechos morales, pero no concluye que la terminología moral sea confusa y prescindible como lo es la terminología religiosa, de acuerdo con un ateo. La gente está en favor de algunas cosas y en contra de otras y aquello respecto de lo cual está a favor o en contra tiene relación con lo que decide hacer. Una decisión práctica requiere un tipo especial de razonamiento que no es simplemente acerca de qué creer sino que concierne a qué hacer. La terminología moral es necesaria

para expresar las consideraciones que son relevantes al razonamiento práctico y la argumentación.

2. *El emotivismo y el argumento de la pregunta pendiente*

Frecuentemente, los emotivistas invocan el argumento de la pregunta pendiente contra el naturalismo ético. El emotivista está inclinado a suponer que la pregunta acerca de si algo que tiene una característica natural dada es incorrecto, siempre será una pregunta pendiente, pues cree que la palabra "incorrecto" conlleva una fuerza emotiva de la cual carece una expresión descriptiva tal como "causa sufrimiento a los seres humanos". El emotivista argumenta que, cualquiera que sea la característica *C* de que se trate, creer que algo es *C* no equivale todavía a tener una actitud de desaprobación, mientras que pensar que algo es incorrecto ya es estar en contra de ese algo. El emotivista argumenta que estar de acuerdo en que algo es *C* es, hasta ese momento, dejar pendiente que se pueda pensar que es incorrecto.

Aunque los emotivistas usan a menudo el argumento de la pregunta pendiente de este modo, se apoyan en algo que va más allá de los principios básicos del emotivismo. Existe además una suposición oculta que debe sacarse a la luz, a saber, la suposición de que los valores no son universales. Supóngase que los valores son universales. En particular, supóngase que siempre que alguien crea que una acción causa sufrimientos a los seres humanos, inmediatamente la desaprueba. Entonces fracasa el argumento de la pregunta pendiente. Pues, en ese caso, creer que una acción causa sufrimiento a los seres humanos es desaprobarla inmediatamente y en consecuencia, dado el emotivismo, es creer que es incorrecta. La pregunta acerca de si una acción que causa sufrimiento a los seres humanos es incorrecta ya no es una pregunta pendiente. La brecha entre "causa sufrimiento a los seres humanos" y "es incorrecto" queda cerrada.

Decir que hay valores universales es compatible con el emotivismo. Tales valores podrían ser innatos, "instalados" [*wired in*] genéticamente en virtud de la evolución. La gente que hubiera nacido sin los valores de sociabilidad y autopreservación incorporados podría ser derrotada en la competencia evolutiva por otros que hubieran nacido con tales valores. El filósofo escocés David Hume (1711–1776) argumenta que la gente está construida de modo tal que siente cierta débil simpatía hacia otras personas. Según Hume, en la medida en que tus propios

intereses no estén en juego, tú estarás automáticamente en favor de lo que es bueno para otra gente. Si Hume está en lo correcto, no puede ser una pregunta pendiente que estés o no a favor de hacer algo que beneficiará a algunas personas y no perjudicará a nadie. Pues pensar que algo lograría tal cosa es estar en favor de ello, de acuerdo con Hume, y eso, según el emotivismo, sería pensar que se trata de algo bueno.

Si se acepta el emotivismo, la medida en la que hay una brecha entre *es* y *debería*, se torna una pregunta psicológica: "¿Qué valores son parte de la naturaleza humana?" Es concebible que, dada la naturaleza humana tal como es, el pensamiento moral siempre descanse en ciertos valores universales básicos. Si esto es verdadero, el emotivismo puede ponerse en conjunción con una teoría ética fuertemente no escéptica y no relativista. Por una parte, siempre es posible que la naturaleza humana sea más variable y que las diferentes personas tengan diferentes valores y principios, y esta variación no implica irracionalidad. Si esto es así, el emotivismo se vuelve una teoría ética más escéptica, más relativista.

El emotivismo dice que los juicios morales expresan actitudes a favor y en contra de las cosas. Esto puede ser parte de una teoría no relativista si se combina con el punto de vista según el cual existe uniformidad en las actitudes humanas básicas. Si el emotivismo se combina con el punto de vista según el cual no existe tal uniformidad, se torna parte de una teoría más relativista. Pero el relativismo no deriva tanto del emotivismo como del punto de vista adoptado con respecto a la naturaleza humana.

Esto significa que el emotivismo no es, en sí mismo, una versión del nihilismo. Es compatible con el emotivismo suponer que hay hechos morales si se supone, además, que hay suficiente uniformidad en la naturaleza humana. Un emotivista puede incluso ser un naturalista ético. Por ejemplo, si un emotivista cree que la aprobación moral y la desaprobación moral derivan de una simpatía universal que sentimos por los demás, entonces puede adoptar una definición naturalista de la incorrección moral: *X* es incorrecto en la medida en que *X* causa sufrimiento a los seres humanos. El argumento de la pregunta pendiente no presentaría ningún obstáculo a esta definición, ya que el emotivista sostendrá que, en virtud de la simpatía universal, no puede ser una pregunta pendiente que algo que causa sufrimiento a los seres humanos sea en esa medida incorrecto. De esta manera, Hume ha sido interpretado como emotivista y también como naturalista ético.

Una diferencia entre el emotivismo y el naturalismo ético emerge sólo si se toman diferentes puntos de vista acerca de la naturaleza humana. Por ejemplo, supongamos que la naturaleza humana no impone constricciones reales sobre los valores básicos y que el razonamiento práctico no crea nuevos valores, sino que únicamente permite realizar tus valores a la luz de cualquier información que tengas. Supóngase, entonces, que no existen valores universales que sea irracional no tener. En otras palabras, supóngase que la razón práctica siempre es y debe ser la esclava de la pasión. Esto sería aceptar el tipo de concepción del razonamiento práctico que encontramos en Aristóteles (384–322 a. C.) o en Hume.

La idea es que uno empieza con ciertos fines, cosas que se favorecen o que se quieren. El razonamiento teórico dice, entonces, que para lograr uno de esos fines se necesita algo más. Esto podría ser algo que uno puede hacer o algo que podría ocurrir. El razonamiento práctico conduce, entonces, a tomar esa otra cosa como un fin intermedio: algo que se favorece o que se quiere pues conducirá al fin originario. Según Hume y Aristóteles, el razonamiento práctico es siempre un razonamiento de medios a fines de este tipo, uno puede llegar a querer algo en tanto medio para algo que ya quería. Pero tal razonamiento puede no tener efecto sobre los fines últimos. No se puede razonar acerca de tener algo como un nuevo fin último, dado que siempre se razona de los fines a las cosas que son medios para esos fines.

No es obvio que esta explicación del razonamiento práctico sea correcta; veremos que hay quienes la rechazan. Pero es una explicación creíble y si se acepta esa explicación del razonamiento práctico y si no hay constricciones interesantes sobre los fines últimos, entonces el emotivismo tiene que ser una teoría relativista y escéptica. El emotivismo, en ese caso, no puede involucrar al naturalismo ético; en realidad, los emotivistas dirán entonces que el argumento de la pregunta pendiente muestra que hay una falacia naturalista.

El naturalista ético identifica que una acción sea incorrecta con que tenga alguna característica natural. Puede decir, por ejemplo, que un acto es incorrecto porque causa sufrimiento a los seres humanos. Ahora bien, es verdad que si los fines últimos de alguien son tales que él está en contra de cualquier acto que cause sufrimiento a los seres humanos, no pensará que es una pregunta pendiente la de si un acto que causa sufrimiento a los seres humanos es, por ello, incorrecto. Pero estamos suponiendo, por el momento, que habrá también gente con diferentes fines últimos, a la cual no le importa

si los actos causan sufrimiento a los seres humanos. Tal gente pensará que (al menos) es una pregunta pendiente si un acto que causa sufrimiento humano es, por ello, incorrecto. Esta gente será capaz de pensar que un acto causa sufrimiento humano sin desaprobarlo y, dado que estamos suponiendo que no es irracional por parte de ellos dejar, de esta manera, de desaprobar hechos que causan sufrimiento humano, de hecho será para ellos una pregunta pendiente si un acto que causa sufrimiento humano es o no, por ello, incorrecto. Así, el naturalista ético en cuestión no habrá captado el modo en que esa gente usa el término "incorrecto".

El naturalista ético puede contestar que no está tratando de captar el uso general sino únicamente su propio uso. Él define sus términos tal como lo hace pues desaprueba los actos que causan sufrimiento humano, aunque haya otra gente que no desapruebe de manera semejante los actos que causan sufrimiento humano. Pero esto es insatisfactorio. La definición del naturalista ético no funcionará ni siquiera cuando se aplique a su propio uso. Supongamos que discute con alguien que no desaprueba los actos que causan sufrimiento humano. Esta otra persona dice que estas cosas *no son incorrectas*. "¡Ah!", responde el naturalista ético, "tal como yo uso las palabras, estas cosas son incorrectas, pues defino 'incorrecto' en términos de 'causa sufrimiento humano' ". Y, por supuesto, el naturalista ético puede usar las palabras del modo en que quiera. Pero eso no pondrá fin al desacuerdo entre él y la otra persona; y el naturalista ético querrá ser capaz de expresar lo que está en disputa en su desacuerdo. ¿Pero cómo? Si su terminología es puramente descriptiva y fáctica, no puede haber modo de expresar el desacuerdo, pues ambos pueden estar de acuerdo acerca de todos los hechos relevantes. No llevará a ninguna parte decir que en tal caso no existe un desacuerdo real, pues existe aún un desacuerdo en la actitud. Por supuesto el naturalista ético puede conceder que se *describa* ese desacuerdo en la actitud. Puede decir que está en contra de algo de lo que la otra persona no está en contra. Pero no puede permitir que ese desacuerdo se *exprese* a menos que en su lenguaje no solamente permita maneras de hablar *acerca* de actitudes, sino maneras de expresar actitudes. Y una vez que permite un modo de expresar actitudes, está usando una terminología que no puede ser definida puramente en términos descriptivos, dada nuestra suposición según la cual la naturaleza humana no es tan uniforme como para que todos tengamos los mismos valores (y dado que la razón es la esclava de las

pasiones). Dada esta suposición, el emotivismo es, en consecuencia, mucho más verosímil que el naturalismo ético.

3. *Ventajas del emotivismo*

Permítaseme ahora repasar brevemente lo que he estado diciendo. El emotivismo es la teoría según la cual tus juicios morales expresan tus actitudes. En otras palabras, tus creencias morales (si es correcto hablar de "creencias" en conexión con esto) no son cognoscitivas, sino que son en sí mismas actitudes a favor o en contra de algo. Según el emotivismo, "Creo que este acto es incorrecto" significa aproximadamente "Desapruebo este acto". Pero "Este acto es incorrecto", por sí misma, no puede ser equiparada a ninguna expresión puramente naturalista tal como "este acto causa sufrimiento humano" a menos que haya una ley natural según la cual pensar que un acto causa sufrimiento humano sea estar en contra de él, sea desaprobarlo.

El emotivismo tiene un buen número de cosas a su favor. En primer lugar, ofrece una explicación verosímil del desacuerdo moral. Tal desacuerdo es un tipo de desacuerdo en la actitud: una persona está a favor de algo y otra está en contra. Además, ciertos desacuerdos morales son insolubles; las partes en el desacuerdo pueden argumentar y argumentar —bastante racionalmente y aun calmadamente— sin alcanzar el acuerdo. El emotivista puede explicar esto suponiendo que las partes en la disputa tienen diferentes valores básicos. En última instancia, quieren cosas diferentes y ninguno de ellos precisa ser irracional o estar confundido por querer lo que quiere.

En segundo lugar, el emotivismo da cuenta de la importancia del problema moral en nuestras vidas y del modo en que las creencias morales están vinculadas a nuestros actos. Estar en favor de algo es querer que eso ocurra. Estar en contra de algo es querer que eso no ocurra. Y lo que quieres influye en lo que haces.

Pensar que debes hacer algo es estar motivado a hacerlo. Pensar que sería incorrecto hacer algo es estar motivado a no hacerlo. El emotivismo puede explicar esto. El naturalismo también podría explicarlo pero sólo suponiendo una ley natural según la cual, digamos, todos desaprueben los actos que causan sufrimiento humano. La explicación naturalista es menos segura, *ad hoc*, y no tan fundamental como la explicación emotivista que no supone una ley natural así.

Éstas son las ventajas del emotivismo: su explicación del desacuerdo moral y su énfasis en la conexión entre las creencias morales, las

pasiones y los actos. Volvamos ahora a las posibles objeciones y a los problemas.

4. *La verdad en la ética*

He mencionado una objeción en el capítulo anterior. ¿Cómo puede dar cuenta el emotivismo del hecho de que llamamos verdaderos o falsos a los juicios morales? Los sentimientos y las pasiones no pueden ser verdaderos o falsos; y, si ciertas aseveraciones simplemente expresan sentimientos o pasiones, no es fácil ver cómo son verdaderas o falsas. Puede ser verdad que me guste algo, que esté a favor de algo; pero gustar, ¿puede ser verdadero? Si tú dices que el secuestro político es incorrecto, puedo estar de acuerdo inteligiblemente diciendo "Eso es verdad". Pero supongamos que tú expresas simplemente tus sentimientos. Dices "¡Bah!" o "¡Ah muchacho!" o "¡Uh!". Entonces, sería inapropiado que yo concordara diciendo "¡Es verdad!" De manera similar, si fueras a emitir la orden "¡Capturen a esos secuestradores!" sería ridículo que yo expresara mi concordancia con tus sentimientos diciendo "¡Eso es verdad!".

Un emotivista *podría* replicar diciendo que al expresar mi concordancia con tus juicios morales diciendo "¡Es verdad!" estoy abusando del término "verdad". Hablando estrictamente, él podría decir que los juicios morales no son ni verdaderos ni falsos, pues expresan actitudes que no pueden ser ni verdaderas ni falsas. Aunque ésta es una réplica posible, está en colisión con una máxima o regla de oro de los filósofos, es decir, la máxima según la cual, si una teoría filosófica está en conflicto con las maneras ordinarias de pensar y de hablar, se debe examinar muy estrechamente la argumentación de tal teoría, pues probablemente algo anda mal.

Éste no es un punto concluyente. Un ateo puede llegar a conclusiones que estén en conflicto con los modos ordinarios de pensar y de hablar, y puede no obstante sentir, quizá correctamente, que son los modos ordinarios de pensar y de hablar los incorrectos antes que su teoría. Lo mismo *podría* ser verdadero para un emotivista que niega que los hechos morales sean verdaderos o falsos. Sin embargo, un conflicto con las maneras ordinarias de pensar y de hablar puede indicar que algo no anda bien, por lo que es útil considerar una línea alternativa que podría tomar un emotivista.

El emotivista podría concordar en que los juicios morales pueden considerarse, con propiedad, verdaderos o falsos. Puede continuar di-

ciendo que lo que esto muestra es que ciertas actitudes y sentimientos pueden considerarse, con propiedad, verdaderos o falsos. Inicialmente esto parece extraño, podría continuar diciendo, pues interpretamos erróneamente lo que implica decir que algo es verdadero o falso. Supongamos que decir que algo es verdadero es decir que corresponde a los hechos. Dado que no vemos cómo corresponden a los hechos los sentimientos y las actitudes, no vemos cómo puede decirse con propiedad que son verdaderos o falsos. Pero esto ocurre porque tenemos un cuadro engañoso de la verdad como correspondencia con los hechos.

Consideremos una verdad de la aritmética: "2 + 2 = 4". ¿Corresponde esto con los hechos? Quizá, ¿pero qué hechos? Si puede decirse que corresponde a algún hecho, éste debe ser el hecho de que 2+2 = 4, si eso se considera propiamente un hecho.

"La nieve es blanca" es verdadera si y sólo si esa aseveración corresponde a los hechos. En otras palabras, la aseveración es verdadera si y sólo si es un hecho que la nieve es blanca. "La nieve es blanca" es verdadera si y sólo si la nieve es blanca, "2 + 2 = 4" es verdadera si y sólo si 2 + 2 = 4. "*S*" es verdadera si y sólo si *S*. Lo que significa decir que la verdad es correspondencia con los hechos es que todos esos principios valen, independientemente de la oración indicativa que reemplace a *S*. Reconocer eso es reconocer qué significa "verdadero". El principio en cuestión vale, ya sea que la oración que reemplaza a "*S*" se use o no para expresar una creencia fáctica o aritmética; vale aunque esta oración exprese una actitud moral. "El incesto es incorrecto" es verdadera si y sólo si el incesto es incorrecto.

Si dices que el incesto es incorrecto y yo digo "¡Es verdad!", simplemente estoy concordando contigo. No estoy diciendo que tu observación corresponda a los hechos en algún sentido fuerte de "hechos". Sólo estoy diciendo esto en un sentido más débil según el cual, cuando digo que es un hecho que el incesto es incorrecto, estoy diciendo únicamente que el incesto es incorrecto.

En otras palabras, el emotivista puede dar cuenta de nuestra tendencia a usar las palabras "verdadero" o "falso" respecto de nuestros juicios morales invocando una teoría de la verdad según la cual "es verdad que *S*" no significa nada más que "*S*". Por lo tanto, quizá hay una buena réplica a la primera objeción al emotivismo.

Al adoptar esta línea argumentativa puede parecer que el emotivista ha abandonado el nihilismo. Como se recordará, el nihilismo es la doctrina según la cual no hay hechos morales ni verdades morales ni

conocimiento moral. Lo que hemos visto es que el emotivista puede argumentar que, en el sentido ordinario de "hecho" y "verdad", hay hechos morales y hay verdades morales. Pero no puede argumentar de una manera similar que hay conocimiento moral, dado que estos "hechos morales" y "verdades morales" no parecen ayudar en la explicación de por qué alguien observa lo que observa, ni pueden reducirse de ninguna manera obvia a hechos y verdades que sean de ayuda para explicar por qué alguien observa lo que observa. El emotivista puede estar de acuerdo en que hay hechos y verdades morales, según cierta manera de hablar, aun según la manera ordinaria de hablar; pero, no obstante, no supondrá que estos hechos y verdades son parte del orden de la naturaleza en el sentido de que la explicación científica de la naturaleza es incompleta si se las excluye, ni supondrá que puede haber conocimiento de hechos de este tipo. Por lo tanto, hay un sentido en el que un emotivista que toma esta línea argumentativa retiene su nihilismo.

Pero, de acuerdo con esto, la primera objeción al emotivismo no ha sido respondida por completo, dado que ordinariamente sí pensamos y hablamos como si a veces supiéramos que deberíamos o no deberíamos hacer distintas cosas. Aquí, quizá, el emotivista puede llegar a decir que nuestros modos ordinarios de hablar y de pensar están simplemente equivocados.

5. *El razonamiento moral*

Consideremos ahora una segunda objeción, la cual pregunta si el emotivismo puede dar cuenta del razonamiento moral. Evidentemente, existe una cosa tal como el razonamiento moral; y algunas veces adoptamos una opinión moral particular —si es que "opinión" es la palabra adecuada— como consecuencia del tal razonamiento. Además nos suena apropiado preguntar por las razones de alguien para sus opiniones morales. Si alguien dice que es bueno que la Unión de Contribuyentes de Oregon haya secuestrado a Sally Jones, le preguntaremos por qué supone que lo es. Nos sonaría extraño que dijese que no tenía razones; que ésa es sólo la manera en que se siente acerca de ello. Eso nos sonaría extraño, pues que diga que el secuestro fue algo bueno indica o incluso implica que (él piensa que) hay razones relevantes para suponer que el secuestro es algo bueno.

El razonamiento no es relevante de la misma manera respecto de los sentimientos. Tú no razonas para que algo te guste o te deje de gustar.

A menudo es verdad que simplemente te gusta algo, o no te gusta algo, sin tener una razón. Por supuesto, algo puede gustarte en ciertos aspectos y disgustarte en otros. Pero esto no es igual a que te guste por ciertas razones o como resultado de un argumento.

En esto existe una diferencia entre la ética y la estética. Existe el razonamiento moral en un sentido en el que no existe el razonamiento estético. No es que no haya nada que pueda denominarse razonamiento estético, sino que la situación en estética es diferente de la situación en ética. Estrictamente hablando, lo que podría denominarse razonamiento estético realmente no es razonamiento en absoluto. Vale para señalar ciertos aspectos del trabajo artístico de modo que alguien los aprecie.

Algo semejante puede también ocurrir en ética. Señalas algunos aspectos de una situación para que alguien los aprecie en forma más completa; argumentando en favor del vegetarianismo, por ejemplo, puedes describir en detalle cómo se trata a los animales antes de matarlos. Pero en ética hay algo más: apelación a principios. Hay principios morales y ellos cumplen un papel en la argumentación moral. Si alguien formula un juicio moral, nos suena apropiado preguntarle cuáles son los principios morales, de una manera que nos parece menos apropiada cuando alguien formula un juicio estético. No es apropiado preguntar a alguien que admira un cuadro cuál es su principio. Y no está claro que el emotivista pueda dar cuenta de la relevancia del razonamiento y la apelación a principios para los juicios morales.

El emotivista sostiene que los juicios morales son expresiones de sentimientos. Pero las expresiones de sentimientos no dependen del razonamiento que parte de principios generales, ni requieren que se las defienda apelando a principios, mientras que los juicios morales dependen del razonamiento que parte de principios generales y requieren que se los defienda apelando a principios. En consecuencia, el emotivista parecería equivocarse al tratar los juicios morales como expresiones de sentimientos.

El emotivista puede contestar que "hay sentimientos y sentimientos". Algunos sentimientos, podría estar de acuerdo, simplemente son reacciones, como los gustos y los disgustos. Otros sentimientos, denominados con más propiedad actitudes, podría continuar, son más complejos e implican los principios y los valores básicos de uno. Las creencias morales no pueden indentificarse con cualquier sentimiento, sino sólo con esos sentimientos y actitudes más complejos.

Pero a menos que el emotivista diga más que esto, su punto de vista adquiere un carácter *ad hoc*. Pues nadie hasta ahora ha dicho por qué ciertos sentimientos deberían diferir de otros de este modo. Es oscuro el por qué un sentimiento o una actitud es el tipo de cosa para la que el razonamiento a partir de principios sería relevante. Pero antes de que consideremos qué más debe decir acerca de esto un emotivista, consideremos una tercera objeción que refuerza la segunda.

Algunas veces efectuamos un juicio moral y posteriormente decidimos que era erróneo. Si nuestros juicios morales son simplemente sentimientos, no está claro cómo podrían, de este modo, ser erróneos. En un momento sentiste de un modo; ahora sientes de otro modo. ¿En qué sentido puedes suponer que tu primer sentimiento era erróneo? Hace algunos años te gustaba ponerle azúcar al café, ahora no te gusta. Por otra parte, si alguna vez pensaste que el incesto es incorrecto y ahora piensas que es correcto, supones que antes estabas equivocado. ¿Cómo este cambio de la creencia moral puede ser simplemente un cambio en tus sentimientos?

6. *La escapatoria de Hume*

Ahora bien, Hume, que puede ser o no ser un emotivista, señala una manera de sortear estas objeciones. Hume subraya que muchos sentimientos y pasiones dependen de que tengamos creencias acerca de los objetos de los sentimientos. Si estas creencias están seriamente equivocadas, se dirá también que el sentimiento es equivocado. En otras palabras, transfieres la palabra "equivocado" de las creencias a los sentimientos que dependen de ellos.

Por ejemplo, puedes estar equivocado en haber confiado en alguien. Tus creencias acerca de la confiabilidad y de los sentimientos de tal persona pueden ser erróneos. Confiaste en él porque tenías ciertas creencias acerca de él. Dado que esas creencias eran erróneas —y dado que no hubieras confiado en él si hubieras sabido la verdad— dices que tu confianza es errónea.

Un revolucionario podría sentirse contento de que la Unión de Contribuyentes de Oregon haya secuestrado a Sally Jones, dado que cree que el secuestro conducirá finalmente a la revitalización de la izquierda revolucionaria. Si su creencia es errónea y el secuestro revitaliza la derecha revolucionaria, dirá que estaba equivocado al sentirse contento.

En otras palabras, se dice que los sentimientos son erróneos, o no, con referencia a creencias que pueden o no ser erróneas, de las que dependen los sentimientos. Si los sentimientos morales son del tipo de sentimientos como la confianza y la felicidad, que dependen de creencias, será apropiado decir que los sentimientos morales son erróneos cuando las creencias relevantes son erróneas. Por ejemplo, si el revolucionario hubiera juzgado inicialmente que el secuestro de Jones era algo bueno, en virtud de lo que creía acerca de sus consecuencias, y luego llegó a creer que las consecuencias iban a ser exactamente las opuestas, hubiera supuesto entonces que su juicio moral inicial estaba equivocado, pues las creencias en las que se basaba eran erróneas.

Esto muestra que el emotivista puede dar cuenta de nuestra inclinación y decir que nuestros propios juicios morales pasados eran erróneos y puede, en consecuencia, responder la tercera objeción. Hume también diría que se puede responder de manera similar a la segunda objeción, concerniente a los principios morales y al razonamiento moral. Pues si los juicios morales expresan actitudes que dependen de las creencias acerca de hechos de cierto tipo, y estas creencias implican suposiciones de un tipo general, entonces el razonamiento relevante a esas suposiciones fácticas puede implicar principios generales. Y, así como nos inclinamos a decir que los juicios morales están equivocados si las creencias correspondientes lo están —transfiriendo así la palabra "equivocado" de las creencias a los sentimientos que dependen de ellas—, también nos inclinamos a decir que razonar a partir de principios generales es relevante a la moralidad si es relevante a las creencias correspondientes; en este caso transferimos la relevancia del razonamiento a partir de ciertas creencias, a los sentimientos asociados que dependen de aquellas creencias.

La diferencia entre la ética y la estética, según Hume, es una diferencia en las pasiones y sentimientos relevantes. Hume piensa que nuestros sentimientos morales implican, entre otras cosas, una simpatía general hacia los demás y, consecuentemente, nuestros sentimientos morales a menudo están basados en creencias acerca de las tendencias generales de actos que afectan los intereses de la gente. Aprobamos moralmente un curso de acción si creemos que ese curso de acción se apoya en principios generales que, si se los sigue generalmente, promoverán la felicidad humana y disminuirán la miseria. Ésta no es la única preocupación moral que tenemos, de acuerdo con Hume, sino una de ellas; y, en virtud de esta preocupación, suponemos que los juicios morales implican una apelación a principios generales.

Los principios generales en cuestión son, en primera instancia, principios acerca de la tendencia de ciertos cursos de acción, si la gente se adhiriera generalmente a ellos, a promover la felicidad y disminuir la miseria. El razonamiento relevante efectivamente concierne a los hechos: el problema es si estos cursos de acción mejorarán de hecho el bienestar. Por un principio de cortesía llamaremos a esto razonamiento *moral* y diremos que los principios relevantes son principios morales. Éste es, entonces, el modo en que el emotivismo puede responder a la segunda objeción.

7. *La trivialidad del emotivismo*

Hasta aquí hemos considerado tres objeciones relacionadas con el emotivismo y hemos visto cómo puede contestarlas un emotivista. La primera objeción fue que a menudo se dice que los juicios morales son verdaderos o falsos; la segunda fue que el razonamiento a partir de principios generales es relevante a los juicios morales; la tercera fue que algunas veces pensamos que nuestros juicios morales están equivocados. El emotivista puede contestar la primera objeción apelando a una teoría redundante de la verdad. Puede responder las objeciones segunda y tercera proponiendo la teoría según la cual transferimos a nuestro hablar relativo a los juicios morales lo que estrictamente se aplica a las creencias de las que dependen los juicios morales.

Pero cuando el emotivismo es defendido de este modo parece menos radical de lo que parecería al principio. Entre otras cosas, el emotivismo no es, ni siquiera en sí mismo, una versión del nihilismo ético, dado que en sí mismo es compatible con el naturalismo ético. Diverge del naturalismo ético sólo cuando se formulan ciertas suposiciones. Además, aun dadas esas suposiciones, el emotivista puede conceder que existen los hechos morales y las verdades morales en el sentido ordinario de estos términos, puede incluso conceder que hay algo denominado con propiedad razonamiento práctico y puede concordar en que algunas veces puede decirse correctamente que la gente está equivocada en sus opiniones morales. Podemos comenzar a preguntarnos en qué difiere el emotivismo del simple sentido común.

NOTA ACERCA DE LECTURAS ADICIONALES

Charles Stevenson discute el desacuerdo en la actitud en "The Nature of Ethical Disagreement", cap. I de *Facts and Values*, Yale University Press,

New Haven, Conn., 1963, pp. 1–9. Véase además su análisis acerca de la verdad en las pp. 214–220.

R.M. Hare trata los enunciados morales como imperativos en *The Language of Morals*, Oxford University Press, Oxford, 1952. Hare ofrece una versión del argumento de la pregunta pendiente contra el naturalismo en las pp. 79–93.

Los escritos de David Hume referidos a ética incluyen su *Treatise on Human Nature*, libros II y III; y *An Enquiry Concerning the Principles of Morals.*

Philippa Foot argumenta que el emotivismo y el naturalismo ético no se distinguen fácilmente en "Moral Beliefs", *Proceedings of the Aristotelian Society*, vol. 59, 1958–1959, pp. 83–104. [La traducción de este artículo al castellano aparece en P. Foot, *Las virtudes y los vicios*, trad. Claudia Martínez, Instituto de Investigaciones Filosóficas, UNAM, 1994, pp. 133–153.

Acerca del problema de si existen principios estéticos, véase Arnold Isenberg, "Critical Communication", *Philosophical Review*, vol. 58, 1949, pp. 330–344.

IV

EL EMOTIVISMO COMO TEORÍA DEL OBSERVADOR IDEAL

1. *La teoría del observador ideal*

El emotivismo, que en principio parece un punto de vista nihilista muy radical, puede admitir el razonamiento moral, las verdades morales y aun los hechos morales. El emotivismo es menos radical y más trivial de lo que parece. Y esto no es todo. Tal como veremos, el emotivismo no se distingue fácilmente de una versión del naturalismo ético denominada "teoría del observador ideal".

De acuerdo con el emotivismo, las opiniones morales son sentimientos. Sin embargo, esto no significa que no pueda decirse con propiedad que las opiniones morales son erróneas. Pues se puede decir que los sentimientos son erróneos si, por ejemplo, dependen de creencias erróneas acerca de los hechos. El emotivismo admite incluso que algunas veces se puede decir correctamente que los *juicios* morales son erróneos. Ya que un juicio que expresa un sentimiento puede ser en algunos casos llamado erróneo si el sentimiento expresado es erróneo.

Tú dices: "Es una cosa buena que Rudolsky haya sido deportado", expresando tu felicidad, que depende de tu punto de vista según el cual su deportación elimina el último obstáculo para el mejoramiento de las relaciones Este-Oeste. Sin embargo, la deportación logra el efecto opuesto y conduce a un empeoramiento de esas relaciones. Entonces, puede decirse que no sólo estabas equivocado en sentirte feliz, sino también en tu aseveración.

Además, al decir lo que dices, implicas que tu felicidad concerniente a la deportación de Rudolsky es justificada y apropiada, al menos desde cierto punto de vista, dados los hechos; y también puede adscribirse esta implicación a tu aseveración. Tu observación tiene esa

implicación. Esto se conecta con el hecho de que tu observación puede ser llamada errónea; puede decirse que es errónea porque tiene una implicación falsa.

Los juicios evaluativos y los juicios morales se enfocan de una manera similar. Si dices que es bueno que Rudolsky haya sido deportado, implicas que se justifica cierta actitud hacia la deportación, al menos desde cierto punto de vista. Cuando menos implicas que la actitud relevante no depende de una visión errada de los hechos.

Esto indica que, aunque un juicio moral sea simplemente una expresión de sentimiento, involucrará una pretensión, a saber, que la actitud expresada es apropiada y justificada, al menos en la medida en que la actitud no dependa de una visión errada de los hechos. En realidad, la actitud en sí misma implica esa pretensión en el sentido de que no puedes tener tal actitud sin suponer simultáneamente que es apropiada y justificada. Esto indica una versión del naturalismo ético que considera que los juicios morales implican juicios *acerca* de sentimientos.

Dado que una actitud es apropiada y justificada en el sentido relevante únicamente si no es errónea —en otras palabras, sólo si no se basa en una visión errada de los hechos—, podemos considerar el siguiente análisis naturalista como una primera aproximación:

> *X* es incorrecto *si y sólo si* desaprobaríamos *X* si conociéramos todos los hechos relevantes y no tuviéramos creencias erróneas concernientes a la cuestión.

Lo mejor sería añadir de inmediato "si apreciáramos completamente todos los hechos", ya que una actitud puede ser incorrecta no sólo cuando se basa en información falsa, sino también cuando descansa en una apreciación impropia de los hechos, de modo que, por ejemplo, algunos de los hechos no se registran. En consecuencia, nuestro análisis se convierte, en una segunda aproximación, en:

> *X* es incorrecto *si y sólo si* desaprobaríamos *X* si conociéramos y apreciáramos completamente todos los hechos relevantes y no tuviéramos creencias erróneas concernientes a la cuestión.

Pero esto no es suficiente. No toda desaprobación es desaprobación moral. Al decir que algo es *moralmente* incorrecto se implica además que tu desaprobación no está basada únicamente en tus propios estrechos intereses, sino que también sería una actitud apropiada desde un

punto de vista completamente desinteresado e imparcial. Implicas que tu desaprobación no deriva únicamente del hecho de que tus propios intereses se vean afectados, y también que tal desaprobación sería apropiada incluso si no tuvieras conexión con la gente involucrada. Si tu actitud deriva únicamente de tus propios intereses y no es sostenida por una óptica imparcial, es injustificada y errónea en tanto pretende ser una actitud moral.

Consideremos a los propietarios de esclavos que aprueban la esclavitud. Es probable que su aprobación descanse, en parte, en ciertas creencias falsas acerca de quienes son esclavos. Probablemente también descanse, en parte, en que no logran apreciar totalmente en qué consiste ser esclavo. Pero, más que nada, casi ciertamente descansa en que no adoptan una óptica desinteresada. Si los propietarios de esclavos consideraran cómo sería su situación si ellos mismos fuesen esclavos, entonces presumiblemente no sentirían aprobación moral respecto de la esclavitud, aunque en la realidad continuaran, por supuesto, estando en favor de que siguiera existiendo la institución con la que se benefician.

Ahora bien, si los propietarios de esclavos pretenden que la esclavitud es moralmente defendible, no pueden aceptar que su aprobación de la misma deriva enteramente del hecho de que están en una posición ventajosa. No podrían defender de manera inteligible su afirmación diciendo simplemente que se percatan de lo que es ser esclavo pero que, como no lo son, no les importa. Ésta es una actitud perfectamente consistente, por supuesto, pero no es una actitud moral. Los propietarios de esclavos pueden seguir sintiendo aprobación moral por la esclavitud sólo en la medida en que sean capaces de convencerse, por ejemplo, de que los esclavos viven mejor en esclavitud, y de que un observador imparcial que conociera todos los hechos también la aprobaría.

De acuerdo con nuestra hipótesis, la esclavitud es incorrecta si y sólo si un observador imparcial desinteresado que conociera todos los hechos y estuviera claramente consciente de ellos y los apreciara, la desaprobara. Los propietarios de esclavos no la desaprobarían, pero sólo porque no serían desinteresados; por lo tanto, su no desaprobación no cuenta.

El tipo de análisis al que hemos llegado se conoce también como teoría del espectador imparcial o teoría del observador ideal. De acuerdo con este análisis, algo es incorrecto si y sólo si un espectador imparcial o un observador ideal lo desaprobara; el observador ideal se

define como una persona desinteresada, bien informada, claramente consciente de los hechos relevantes, etcétera.

De ser correcto, este análisis proporciona una reducción naturalista de los hechos morales. Hume, quien acepta esta óptica, la compara con la teoría análoga del color que hemos discutido en el capítulo 2 y que dice que algo es rojo, por ejemplo, si luciera rojo ante un perceptor normal en condiciones luminosas estándar. Para este punto de vista, los colores son propiedades disposicionales, como también lo son las propiedades morales. Una propiedad disposicional debe ser definida en términos de una propiedad que acaezca concurrentemente. Por ejemplo, la propiedad disposicional de ser soluble en agua se define en términos de la prioridad concurrente de disolverse en agua. Una sustancia es soluble en agua si se disuelve cuando se coloca en agua. De manera similar, un objeto es rojo si luce rojo en ciertas condiciones.

Una sustancia puede ser soluble en agua aunque nunca haya sido colocada en agua. De manera similar, un objeto puede ser rojo aunque esté en una habitación oscura. Cuando cae un árbol en el bosque y no hay nadie allí, el árbol produce un ruido muy fuerte en el sentido disposicional, ya que si alguien hubiese estado allí habría escuchado un fuerte ruido.

La propiedad concurrente respecto de los colores es que un objeto parezca tener cierto color. Podríamos hablar de sensaciones de color. Un objeto es rojo si transmite a los observadores la sensación de color rojo, así como algo es caliente si produce una sensación de calor a los observadores. Y algo es incorrecto si produce "sensaciones de desaprobación" en ciertos observadores. La incorrección es una disposición a causar desaprobación.

2. *El relativismo moral y la teoría del observador ideal*

No necesitamos suponer que todo observador ideal tendría las mismas reacciones: algunos podrían desaprobar una situación dada, otros podrían ser indiferentes y otros incluso aprobarla. Las reacciones de un observador dependerán del tipo de educación que haya tenido. En ese caso, tendríamos que dar cabida a un tipo de relativismo moral. Tendríamos que distinguir, digamos, la incorrección-Nueva York, la incorrección-Pekín y la incorrección-Moscú, así como distinguimos la solubilidad en agua de la solubilidad en aceite. No podemos hablar simplemente de solubilidad porque diferentes sustancias son solubles en líquidos diferentes. De modo similar, puede ser que no podamos

hablar simplemente de incorrección, ya que observadores ideales con diferente educación podrían desaprobar diferentes cosas. Tendríamos que decir que algo era incorrecto en Nueva York pero no incorrecto en Pekín si lo desaprobaran los observadores ideales criados en Nueva York pero no los observadores ideales criados en Pekín.

Pero podríamos decir que algo es simplemente correcto o incorrecto si el contexto hiciese claro que sólo los observadores ideales de un tipo dado eran relevantes. Por ejemplo, si todos nosotros nos hubiéramos criado en Nueva York, podríamos decir que algo era simplemente incorrecto queriendo decir que era incorrecto en Nueva York. De modo similar, cuando está claro que sólo es relevante la solubilidad en agua, podemos decir que una sustancia es soluble, comprendiéndose que eso significa que es soluble en agua.

Nuestro uso de la expresión "está a la moda" ilustra este tipo de relativismo. Decir que algo está a la moda es decir que le parecería a la moda a un observador apropiado. Pero el hecho de que algo te parezca a la moda dependerá de los tipos de cosas que estés acostumbrado a ver. Las modas de otra época o de otro lugar a menudo parecen grotescas. Decir que algo "está a la moda", en consecuencia, es efectuar un juicio altamente relativista. Por lo tanto, *X* está a la moda en tal y tal momento y lugar si y sólo si *X* le pareciera a la moda a un observador normal de la moda de ese momento y lugar. Los juicios morales también pueden ser relativistas, aunque quizá no en tal medida.

La teoría del observador ideal no necesita ser una teoría relativista. No hay compromiso de ninguno de los dos. En la teoría del observador ideal, el que el relativismo sea verdadero depende de hechos de la psicología humana. En el caso de los colores, por ejemplo, el relativismo *no* es verdadero. No todos los que perciben tienen las mismas experiencias de color en condiciones estándar, ya que algunos de quienes perciben son daltónicos. Éstos pueden ser clasificados como defectuosos ya que, en primer lugar, sus reacciones no son tan sensibles como las de quienes perciben normalmente (los daltónicos realizan menos discriminaciones de color) y, en segundo lugar, los percipientes normales que están en condiciones de realizar la mayoría de las discriminaciones de color concuerdan en sus experiencias de color. Lo que es verdadero para el color también *podría* resultar verdadero para las propiedades morales. Ya que es *concebible* que podamos explicar los desacuerdos entre los observadores morales aparentemente ideales suponiendo que algunos otros observadores ideales tienen

algún tipo de ceguera moral que resulta de algún tipo de defecto genético o de alguna falla en su educación, por ejemplo, falta de un afecto adecuado en sus primeros años. Para afirmar esta pretensión debería mostrarse que este grupo de observadores aparentemente ideales no realiza todas las discriminaciones morales que realizan otros observadores ideales, aunque los otros observadores sí realizan todas las discriminaciones morales que hace el primer grupo. Además, debería mostrarse que los otros observadores ideales tienen, en la mayoría de los casos, las mismas reacciones morales.

No es del todo implausible que algo así pudiera mostrarse. Los estudios psicológicos concernientes al desarrollo moral de los niños indican que existe un número de estadios en su desarrollo, estadios que muestran una sofistificación creciente. No todos alcanzan los estadios finales de desarrollo. Podríamos querer definir la corrección o incorrección únicamente en términos de las reacciones de aquellos que hayan alcanzado el estadio de desarrollo moral más alto y tratar a los otros como si poseyeran un tipo de ceguera moral o alguna otra deficiencia.

3. *El emotivismo* versus *la teoría del observador ideal*

Ésta es, pues, la teoría del observador ideal. Llegamos a esta teoría por vía del emotivismo. El emotivismo, que pretende ser una versión del nihilismo moderado, conduce a la teoría del observador ideal, que es una versión del naturalismo ético. Ya que el emotivismo dice que los juicios morales expresan actitudes y que, al expresar el tipo de actitud relevante, el hablante implica que la actitud es apropiada y justificada al menos desde cierto punto de vista.

Que ésta sea una implicación de lo que dices o solamente una implicación de haberlo dicho es una cuestión sutil. Comparémosla con un caso análogo: cuando dices que Rudolsky fue deportado, implicas que está justificada la creencia de que Rudolsky fue deportado; pero no *dices* que esta creencia esté justificada; tampoco *lo que dices* implica esto, aunque *el que lo digas* lo hace. De manera similar, aunque al realizar un juicio moral implicas que cierta actitud está justificada, de esto no se sigue que podamos analizar tu juicio como si *dijese* que se justifica esa actitud.

Existen dos posibilidades: la teoría del observador ideal puede ser tratada como un análisis de lo que estás diciendo cuando efectúas un juicio moral, o bien puede ser tratada como un análisis de algo que

tú implicas al efectuar el juicio. Estas posibilidades no se distinguen fácilmente en la práctica, aunque son distinguibles en teoría.

Existe un último argumento no concluyente para suponer que la teoría del observador es un análisis de lo que estás diciendo y no un análisis de algo que implicas al efectuar un juicio moral. El argumento puede ser visto como un argumento contra el emotivista, que cuestiona la idea de que los juicios morales simplemente expresan actitudes a favor o en contra de algo. El argumento es que la relación entre tus juicios morales y aquello respecto a lo cual estás a favor o en contra es más indirecta de lo que el emotivista está dispuesto a conceder. Puedes pensar que deberías hacer algo sin querer hacerlo. En realidad, éste es el núcleo de cierto tipo de dificultad moral: advertir que debes hacer algo que no quieres hacer. De manera similar, puedes advertir que otro debería hacer algo que tú esperas que no haga. En general, hay muchos casos en los que aquello de lo que estás en favor difiere de lo que piensas que debería ser el caso. En consecuencia, los emotivistas están equivocados (así continúa el argumento) cuando identifican el pensamiento según el cual algo debería ser el caso, con estar a favor de ello. La teoría del observador ideal concebida como una teoría acerca del *contenido* de un juicio moral puede explicar fácilmente estos hechos recalcitrantes. Ya que puedes creer que, en condiciones de imparcialidad, tendrías una actitud particular aunque no tengas esa actitud en las condiciones vigentes. De modo análogo, efectúas con facilidad juicios morales acerca de situaciones hipotéticas. Pero, ¿qué te importa lo que pase en una situación hipotética? Quizá sea cierto que si el médico *fuera* a despedazar al paciente de la habitación 306 y fuera a distribuir sus órganos entre los cinco pacientes moribundos, tú desaprobar*ías* que lo hiciera. El emotivismo no es la teoría según la cual algo es incorrecto si en ciertas condiciones tú lo desaprobar*ías*. Ésa es la teoría del observador ideal tomada como un análisis de lo que dice un juicio moral. De acuerdo con el emotivismo, cuando en este momento digo que el médico haría algo incorrecto, estoy expresando en este momento mi desaprobación actual de ese acto hipotético. Pero, ¿qué puede significar esto?

Aunque ésta parece ser una objeción poderosa contra el emotivismo, puede ser enfrentada de una manera tan sencilla como las objeciones anteriores invocando la estrategia de "hay sentimientos y sentimientos". El emotivista dirá que, cuando piensas que deberías hacer algo pero no quieres hacerlo, favoreces que se haga pero también favoreces que no se haga; tienes una actitud moral a favor de

hacerlo y una actitud más fuerte, pero no moral, a favor de no hacerlo. De manera similar, si piensas que deberías hacer algo pero no te preocupas, el emotivista dirá que te preocupas y que no te preocupas. Tu actitud moral favorece que hagas la cosa en cuestión, pero esa actitud es débil y en consecuencia es superada por tu inercia general. Como en los casos hipotéticos, el emotivista sostendrá que te preocupas moralmente por esos casos aunque en algún otro sentido no te preocupas. Distinguiendo así las actitudes morales de otras actitudes, el emotivista puede hacer frente a la objeción y, consecuentemente, socavar el argumento precedente tomando la teoría del observador ideal como un análisis de lo que dices cuando efectúas un juicio moral.

4. *La amenaza de circularidad*

Esta respuesta, no obstante, representa otro paso en la trivialización del emotivismo. El emotivista puede muy bien dar un paso más y simplemente *estipular* que las creencias morales son actitudes. En otras palabras, puede muy bien tomar la creencia acerca de que algo es moralmente correcto como estar moralmente en favor de ello: estás a favor de ello al menos en la medida en que piensas que debería hacerse. De este modo, casi puede hacerse que el emotivismo sea verdadero por definición.

Esto es posible porque la palabra "actitud" es muy vaga. Puede decirse que las creencias son actitudes. Por ejemplo, tu actitud respecto de la guerra podría consistir en que piensas que no tendrá lugar por un par de meses. Hay un sentido en el que, trivialmente, las creencias morales son actitudes. Son actitudes en el mismo sentido en el que toda creencia es una actitud.

Si concordamos simplemente en estipular que creer que algo es incorrecto es un modo de estar en contra de ello, entonces el emotivismo difícilmente puede distinguirse del punto de vista trivial según el cual decir que algo es incorrecto es expresar la creencia de que es incorrecto. El emotivismo casi se transforma en la teoría vacua según la cual la creencia de que algo es incorrecto es la creencia de que ello es incorrecto.

Por lo tanto, el emotivismo está amenazado por la trivialidad, a menos que pueda encontrarse algún modo no circular de identificar las actitudes morales. Además, un problema análogo amenaza la teoría del observador ideal. Pues, ¿qué tipo de desaprobación debe sentir el observador ideal para que algo sea incorrecto? Decir, simplemente,

"desaprobación" es inadecuado. El observador ideal podría, por ejemplo, sentir desaprobación estética sin que hubiese nada moralmente incorrecto. Si Osvaldo ejecuta una danza que un observador ideal desaprobaría basándose en fundamentos puramente estéticos, Osvaldo no ha hecho aún nada moralmente incorrecto. Pero, si decimos "desaprobación moral", ¿cómo va esto más allá de "el juicio según el cual algo es incorrecto"? Definir "*X* es incorrecto" como "un observador ideal sentiría desaprobación moral hacia *X*" sería semejante a definir "*X* es incorrecto" como "un observador ideal pensaría que *X* es incorrecto". Y eso sería circular; estaríamos usando el término "incorrecto" para analizar el término "incorrecto".

En consecuencia, quienes proponen tanto el emotivismo como la teoría del observador ideal necesitan especificar un modo no circular de distinguir las actitudes morales de las actitudes no morales si quieren escapar a la trivialidad o a la circularidad. Por otro lado, dada tal especificación, cada una de las posiciones puede ser defendible y no ser fácilmente distinguible de la otra.

Será difícil distinguir el emotivismo de la teoría del observador ideal porque una actitud moral involucra una pretensión; una pretensión que la teoría del observador ideal trata de explicar. No puedes adoptar sinceramente la actitud sin aceptar la pretensión. Tal como lo hemos visto, no sólo las actitudes morales involucran pretensiones de este modo. No puedes estar *contento* porque Rudolsky haya sido deportado, si crees que tu alegría es inapropiada o errónea dados los hechos. En realidad no puedes *creer* que Rudolsky fue deportado si esta creencia no está garantizada por las pruebas. Muchas actitudes involucran la pretensión de que están justificadas desde uno u otro punto de vista. Las actitudes morales involucran la pretensión de que están justificadas desde el punto de vista moral, esto es, desde un punto de vista imparcial.

5. *Los principios morales y el observador ideal*

Un punto final. La teoría del observador ideal fue introducida por analogía con la teoría de las propiedades de color. Pero hay una importante falta de analogía. El percipiente normal o estándar al que nos referimos en la explicación de los colores es considerado la idealización de un percipiente ordinario. Pero el observador ideal en la teoría del observador ideal de las propiedades morales no es una idealización de un juez moral ordinario. Ya que la persona apela esencialmente a

los principios morales en su razonamiento moral, mientras que no hay más razón para que un observador ideal apele a los principios morales que la que habría para que alguien que percibe un color apele a la teoría del espectro para percibir el color o para que un amante de la música deduzca su apreciación de un primer movimiento de los principios de la armonía. El observador ideal tiene todo en su mente y luego reacciona. Esto es muy diferente de la imagen corriente acerca de cómo se alcanza una conclusión moral. En la moralidad, los principios son relevantes de un modo en el que no lo son en la percepción del color o en la apreciación musical. El juicio moral es, según lo suponemos ordinariamente, una cuestión de principio.

Ahora bien, Hume dice, como se recordará, que los así denominados principios morales son en realidad generalizaciones acerca de la utilidad social. Si Hume está en lo cierto, después de todo, el observador ideal hará uso de los principios morales, dado que estamos suponiendo que apela a los hechos. Pero debemos considerar con más cuidado si la respuesta de Hume abarca todas las apelaciones a principios en la moralidad.

Ciertamente, la respuesta de Hume funciona en algunos casos. Algunos principios morales simplemente representan creencias generales acerca de los hechos. Por ejemplo, los principios morales asociados al conservadurismo y al liberalismo a menudo representan diferentes ópticas económicas. Considérese el principio según el cual, en los tiempos de escasez, es incorrecto por parte de una gran corporación tomar ventaja de la situación elevando los precios a modo de obtener grandes utilidades. Que aceptes o no este principio depende en gran medida de tus creencias económicas acerca de los efectos a largo plazo de un mercado relativamente libre *versus* una economía relativamente controlada.

Pero hay otras cuestiones de principio que no se ajustan a este patrón. Consideremos los argumentos concernientes al aborto o acerca de si la gente debería volverse vegetariana. Tales argumentos involucran una apelación a principios generales de una manera diferente. La apelación a principios es una parte esencial de tales argumentos. Cualquiera puede apreciar la fuerza de la pregunta "¿cómo puede justificarse el aborto sin justificar el infanticidio?" o bien "¿cómo puede justificarse que se coman cerdos pero no bebés?" Se nos pregunta "¿qué principio distingue estos casos?"; y ésta parece ser una pregunta legítima. Nos sentimos compelidos a responder tal pregunta si

queremos continuar creyendo que existe una diferencia moral entre los casos.

Hay algo acerca de la moralidad que nos lleva a suponer que los juicios morales deben defenderse mediante la apelación a un principio que distinga los casos aparentemente similares respecto de los cuales se han formulado diferentes veredictos. Ésta es una diferencia importante entre la moralidad y la estética, dado que tú puedes pensar que una melodía es buena, que otra es banal, sin sentirte compelido a encontrar principios estéticos que las distingan.

Esto se relaciona con un punto mencionado anteriomente. El emotivismo y la teoría del observador ideal no son específicamente teorías morales. Pueden también con mucha facilidad ser presentadas como teorías de las propiedades estéticas. De hecho, parecen más apropiadas para la estética que para la moralidad, ya que el juicio estético es más una cuestión de apreciación que de decisión basada en principios.

Sin embargo, esto no significa que se haya mostrado que la teoría del observador ideal no se aplica a la moralidad. Las teorías del observador ideal pueden indicar, correctamente, que no es obviamente relevante que el observador ideal use principios morales. El observador ideal no pretende ser una idealización de alguien que alcanza una decisión moral. La relación entre nuestros juicios y las reacciones del observador ideal es diferente. Lo que se necesita es que la teoría del observador ideal dé cuenta de nuestra apelación a principios. No hay razón teórica por la cual el observador ideal necesite apelar a principios.

Cuando preguntamos qué principio distingue el aborto del infanticidio, estamos preguntando —de acuerdo con la teoría del observador ideal— no por los principios de un observador ideal, sino por las reacciones de un observador ideal. Estamos preguntando por alguna razón para pensar que un observador ideal reaccionará de manera diferente en ambos casos.

Quien juzga que el aborto es moralmente correcto pero que el infanticidio no lo es, puede efectuar este juicio porque no es desinteresado, no es imparcial. Puede haber una tendencia a tomar en cuenta las necesidades del infante, dado que se parece más a nosotros, pero a no tener en cuenta las necesidades del feto. Similarmente, los juicios morales racistas reflejarán parcialidad hacia la raza propia; y los que defienden una especie son parciales respecto de su propia especie. Un prejuicio de especie puede dar cuenta de nuestra tendencia a pensar

que es moralmente correcto comer cerdos pero que no lo es comer bebés.

Ahora bien, se supone que un observador ideal es desinteresado. No será parcial en favor de una raza, en favor de un sexo, en favor de una especie o en favor de los infantes sobre los fetos. De acuerdo con la teoría del observador ideal, cuando preguntamos por el principio que justifica el aborto pero no el infanticidio, preguntamos qué hay acerca del aborto y del infanticidio que llevaría al observador ideal a aprobar uno pero no el otro. En consecuencia, en la teoría del observador ideal podemos estar en condiciones de dar cuenta del hecho de que el razonamiento moral apela de manera esencial al principio moral, aun cuando al dar esta explicación tengamos que abandonar al menos parte de nuestra analogía original entre la teoría del observador ideal de las propiedades morales y una explicación natural de las propiedades de color.

6. *Una moraleja*

El emotivismo no es completamente trivial; formula una buena observación. Pensar que algo es bueno, que es lo que hay que hacer, lo que debería ser hecho, etc., son modos de estar en favor de algo; y pensar que algo es malo, que es lo que no hay que hacer, que no debería ser hecho, etc., son modos de estar en contra de algo. Esto no es sólo una estipulación; refleja algo importante acerca de las creencias morales y de los juicios morales. No podríamos *estipular* plausiblemente que pensar que algo es bueno es un modo de estar *contra* algo.

La teoría del observador ideal tampoco es completamente trivial. Explicita el punto de que las creencias morales involucran una pretensión de imparcialidad. Esto se expresa a veces diciendo que los juicios morales son "universalizables": si crees que *X* es incorrecto tienes que creer también que todo lo que sea relevantemente semejante a *X* también será incorrecto, independientemente de las personas que estén involucradas, y de cómo estén involucradas. (Por ejemplo, si piensas que una situación dada que te involucra a ti y a alguien más es incorrecta, estás comprometido a suponer que la misma situación sería incorrecta si las posiciones se invirtieran.)

Combinando ambos puntos, vemos que puedes tener creencias morales sólo en la medida en que seas susceptible de un tipo de motivación imparcial. Ya que, si estás en favor de algo, estás en esa medida motivado a tratar de ponerlo de manifiesto, y estar moralmente en favor

de algo es estar en favor de ello de un modo imparcial. En el próximo capítulo comenzaremos a considerar cuál puede ser la fuente de tal motivación imparcial.

NOTA ACERCA DE LECTURAS ADICIONALES

La teoría del observador ideal es defendida por Roderick Firth en "Ethical Absolutism and the Ideal Observer", *Philosophy and Phenomenological Research*, vol. 12, 1952.

Richard B. Brandt discute la teoría en relación con otros puntos de vista en *Ethical Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1959, caps. 7–11.

Jean Piaget describe estadios de desarrollo moral en su libro *El criterio moral en el niño*, trad. Nuria Vidal, Fontanella, Madrid, 1971.

3. LA LEY MORAL

V

SOCIEDAD Y SUPERYÓ

1. *Las reglas morales*

La moralidad, tal como la concebimos corrientemente, involucra múltiples prohibiciones y requerimientos. No debes mentir; no debes robar; no debes violar tus promesas; no debes matar; debes ayudar a quienes lo necesiten; debes respetar a los demás; debes estar agradecido con quienes te han ayudado, especialmente con tus padres; debes ser leal a tus amigos; no debes revelar secretos, etcétera.

Estas reglas componen un tipo de moral, de modo que el razonamiento moral a menudo es semejante al razonamiento legal. Por ejemplo, está prohibido matar. Matar a la gente, al menos. Pero no todo matar es incorrecto. No violas la ley moral si matas en defensa propia. Y, algunas veces, tampoco violas la ley moral si matas a una persona para salvar a otras. He aquí un ejemplo que han discutido los filósofos:

> Vas conduciendo un tranvía y los frenos fallan. Hay cinco personas trabajando en los rieles, vueltos de espaldas. Afortunadamente puedes pasar a un riel lateral si actúas con rapidez. Por desgracia también hay alguien vuelto de espaldas en ese riel. Si pasas tu tranvía al riel lateral, matarás a una persona. Si no lo cambias, matarás a cinco personas.

En tal caso parecería que la ley moral te permitiría pasar el tranvía al riel lateral y matar a una persona antes que a cinco.

Comparemos este caso con el que hemos discutido previamente en el que un médico despedaza y, en consecuencia, mata a un paciente sano a modo de salvar a los otros cinco pacientes. Eso parecía moralmente incorrecto mientras que no parece incorrecto chocar contra el

trabajador del riel lateral, al menos no parece obviamente incorrecto de la manera en que la acción del médico parecía obviamente incorrecta.

Al considerar estos ejemplos y casos similares, nos encontramos inquiriendo por las diferencias que darían cuenta de nuestros juicios intuitivos. Al hacer tal cosa, razonamos como los abogados, o como los jueces. Por ejemplo, al intentar dar cuenta de la diferencia en estos dos casos, podemos llegar a aceptar lo que ha sido denominado el principio del "doble efecto". Según este principio, existe una importante distinción entre lo que te propones, ya sea como uno de tus fines o como un medio para uno de tus fines, y lo que meramente prevés que ocurrirá como una consecuencia de tu acción. Es peor, por ejemplo, pretender dañar a alguien, ya sea como fin o como medio, que pretender algo que sabes que conducirá al daño de alguien. Hacer algo que causará un daño a alguien es muy malo; pero, de acuerdo con el principio del doble efecto, es aún peor pretender tal daño. Ahora bien, el conductor del tranvía no pretende dañar al trabajador del riel lateral. Si el trabajador salta a un costado y se salva, eso no interfiere en el propósito del conductor del tranvía. Si el trabajador resulta lesionado, ése es el *efecto lateral* de lo que hace el conductor del tranvía; no es parte de sus *medios* para salvar a los otros trabajadores. Pero el médico se propone dañar al paciente sano al que despedaza para salvar a los otros cinco pacientes. Si el paciente sano escapara, ello desbarataría el plan del médico. Despedazar al paciente sano es el medio del médico para salvar a los otros pacientes; no es un efecto lateral de lo que el médico hace. En consecuencia, de acuerdo con el principio del doble efecto, es peor salvar a los otros pacientes que atropellar al trabajador del riel lateral, aunque en ambos casos se salven cinco vidas y se pierda una.

La prueba de que nuestra moralidad incorpora principios como el del doble efecto es que tales principios son necesarios para dar cuenta de nuestros juicios morales intuitivos. Por supuesto, que nuestra moralidad realmente involucre estos principios y no otros no puede demostrarse mediante esta breve discusión, la cual, sin embargo, sí debería ser adecuada para indicar que nuestro sistema moral es complejo. El razonamiento moral frecuentemente involucra distinciones sutiles de un tipo que tiene la característica del derecho. Por lo tanto, si la moralidad no es una ilusión, tiene que haber una ley moral o algo que se le parezca.

La idea de que hay algo semejante al derecho moral se refleja en nuestro uso del lenguaje. Hablamos de obligaciones morales, de deberes morales, de derechos morales y de excusas morales; y estas nociones —deber, derecho, obligación y excusa— parecerían tener sentido únicamente en relación con algún tipo de ley. Por ejemplo, no se aplican a juicios estéticos o prudenciales; no hay obligaciones deberes o derechos estéticos o prudenciales. (Es cierto que algunas veces decimos que alguien "se debe a sí mismo" hacer algo dado, pero éste es un modo metafórico de hablar.)

El punto se extiende a palabras del tipo de "debería", "tiene que" y "poder". Aunque estas palabras tienen usos no morales, cada una es usada además en un sentido moral particular y en ese sentido decir, por ejemplo, que deberías hacer algo es decir que hacerlo es tu obligación moral o tu deber. Dado que los juicios acerca de la obligación y el deber sólo tienen sentido en relación con algún tipo de derecho, estos juicios morales "debería" también tienen sentido sólo en relación con algún tipo de derecho.

El punto puede tornarse borroso dada la sistemática ambigüedad de estas palabras. Cada una de estas palabras tiene al menos una cuádruple ambigüedad. Existe, por ejemplo un "debería" de expectativa, como cuando decimos que el tren ya debería de haber llegado. También hay un "debería" de evaluación, como cuando decimos que debería haber más amor en el mundo, sin implicar que alguien en particular tiene la obligación de cuidar que haya más amor en el mundo. Luego hay un "debería" de razones, como cuando decimos que un ladrón debería usar guantes, significando que tiene una razón para hacerlo. Finalmente, está el "debería" moral, como cuando decimos que el ladrón no debería robar. Este "debería" es semejante al "debería" de razones excepto en que se le usa de un modo más restringido para indicar razones morales. Tal como he señalado, estas razones tienen que derivarse de algo semejante a una ley moral. (Se dirá algo más acerca de los sentidos de "debería" en el capítulo 7.)

2. *La sanción moral*

Ahora bien, el rasgo del derecho es que tiene sanciones. Entonces, ¿quién hace cumplir la ley moral? ¿La sociedad? ¿Es la sanción la presión social? La dificultad con esta idea, que por lo demás resulta verosímil, es que la sociedad sólo puede hacer cumplir los principios que son generalmente aceptados por los miembros de la sociedad;

y todos concordamos en que lo que la moralidad dicta y lo que la sociedad dicta no siempre coinciden. Por lo tanto, la ley moral no puede ser simplemente lo que la sociedad impone mediante la presión social.

Pero algo similar a esa idea puede ser correcto. Consideremos cómo se aprende la moralidad. Al principio, "*X* es incorrecto" o "*X* es malo" significa algo semejante a "Mamá y papá no quieren que haga *X*". El concepto que el niño tiene de sus padres es el de seres omniscientes y omnipotentes cuyo amor es crucial para él; el niño trata de complacerlos para que lo amen y no lo castiguen. Pero no siempre está seguro de lo que quieren. Precisa un modo de predecir sus deseos. Por ello usa su imaginación. Trata de imaginar cómo reaccionarían sus padres ante las diferentes cosas que podría hacer. Intenta ponerse en su lugar para ver cómo reaccionaría si fuera ellos. De este modo llega a simular en su imaginación que él es un padre y luego trata de actuar a modo de complacer a su padre imaginario simulado. Para usar la terminología freudiana, el niño desarrolla un "superyó" o un "ideal-de-yo". Internaliza ciertos requerimientos morales. Estos requerimientos no son necesariamente los que sus padres le han inculcado, ya que el superyó es un padre modificado e idealizado y puede ser más estricto o más indulgente que los padres reales del niño, por razones que a los psicoanalistas les parecen interesantes.

Pero nosotros no necesitamos suponer nada acerca de la teoría psicoanalítica. Sean cuales fueren los detalles psicoanalíticos, el mecanismo psicológico en cuestión no es peculiarmente oscuro. Puede ser observado, por ejemplo, cuando un atleta internaliza a un entrenador imaginario. El atleta se habla a sí mismo como lo haría un entrenador, se alienta a jugar en forma más intensa y mejor, castigándose por los errores que comete. El atleta simula ser su propio entrenador. Lo hace para jugar mejor y en una forma más adecuada y, a menudo, es más duro consigo mismo de lo que lo sería el entrenador.

De la misma manera, el niño llega a internalizar principios de moralidad. Habiendo formado un superyó, el niño trata entonces de conservar su amor y de evitar sus castigos. Así, "*X* es incorrecto" llega a significar algo parecido a "*X* es incompatible con las demandas de mi superyó". Por supuesto, el propio niño es su superyó. Él es su propio superyó en la medida en que adopta el papel de sus propios padres. Entonces, al tratar de actuar moralmente; el niño está tratando de obtener su propio amor y respeto. Esto continúa hasta después de

que el niño ha crecido y se ha transformado en un adulto. En tanto trata de actuar moralmente, trata de actuar a modo de conservar el amor y el respeto de aquella parte de sí mismo que desempeña el papel de sus padres imaginarios.

En consecuencia, la moralidad tiene una fuerza social. En primer lugar, la sociedad, actuando a través de los padres, impone una ley moral social. Los niños internalizan esta ley mediante la formación de un superyó en la imagen idealizada de sus padres. Y esto sugiere que los complejos detalles de esta ley moral han de ser explicados históricamente, quizá en términos de función social. La sociedad funciona mejor si la gente en general dice la verdad, mantiene sus promesas, no roba, no mata a los otros, etcétera. En consecuencia, los requerimientos de la ley moral no necesitan ser únicamente requerimientos de razón, en un sentido estrecho. Puede no haber nada *a priori* acerca de la moralidad, desde el punto de vista de un individuo, aunque ciertos requerimientos pueden ser necesarios si la sociedad ha de sobrevivir y estos requerimientos, en cierto sentido, serían necesarios para la moralidad.

Esto no quiere decir que la ley moral sea exactamente la misma ley que impone la sociedad. Antes criticamos esta idea por ser inconsistente con el hecho trivial de que lo que es correcto puede estar en conflicto con lo que la mayoría de los miembros de la sociedad creen que es correcto y, consecuentemente, imponen. Ahora estamos en condiciones de ver que, sin embargo, algo semejante a esa idea puede ser correcto. La moralidad del individuo puede derivarse de una ley que la sociedad impone, sin ser idéntica a tal ley en todos los aspectos. Tu superyó es una versión idealizada de tus padres, quienes en primera instancia son personas que te transmiten las demandas de la sociedad. Las demandas de tus padres pueden no coincidir exactamente con las demandas que otros padres hacen a sus hijos, y además la versión idealizada que tienes de ellos puede diferir de la versión real en distintos modos.

Ahora bien, una persona puede identificar o no con Dios a su padre imaginario. Si lo hace, su moralidad tomará la forma de una moralidad basada en la ley divina. Esto tiene la desventaja de que, si Dios no existe, su moralidad descansará en una ficción. Pero el adulto no religioso parece estar en una posición aún peor. Exista o no exista Dios, su moralidad descansa en una ficción, pues su moralidad descansa en las supuestas órdenes de un padre imaginario.

3. *¿Puede evaluarse la moralidad?*

Es probablemente cierto que los mecanismos psicológicos que conducen al desarrollo del superyó son útiles desde el punto de vista social. Ya que la moralidad es útil desde el punto de vista social y el desarrollo del superyó es lo que hace a la gente actuar moralmente (cuando lo hace). En consecuencia, el superyó es una ficción socialmente útil. Pero es una ficción, una simulación. Ése es el punto. La moralidad descansa en una simulación y una ficción. Decir que la moralidad es socialmente útil no constituye una defensa de la moralidad y de la ficción en la que descansa. Dado el punto de vista moral, tenemos una razón para conservar algo que es socialmente útil. Pero defender al superyó de esta manera constituye un razonamiento circular. Esto no es decir nada más que tenemos una razón moral para preservar la moralidad. Dado que esta razón moral descansa en una ficción, ésta es sólo una razón ficticia.

Pero el hábito que desarrollas de simular que eres un padre idealizado no es, por muchas razones, un hábito fácil de modificar o de manejar, al descansar tal como lo hace en tus anhelos tempranos de amor de tus padres. No hay una manera fácil de abandonar a tu superyó y, en consecuencia, no hay manera fácil de abandonar la moralidad. No es suficiente, por ejemplo, decidir luego de una reflexión psicológica y filósofica que la moralidad es una simulación gigantesca. El superyó no desaparecerá tranquilamente. Representa un hábito de acción que está enclavado en la estructura de tu personalidad, el cual sirve a necesidades psicológicas profundas al permitir que te identifiques con tus padres; un hábito defendido por el mecanismo de la represión, que se hace muy difícil asir.

No obstante, parecería que deberías al menos ser capaz de plantear la pregunta acerca de si es racional ser moral; con otras palabras, si es racional permitirte a ti mismo seguir los dictados de tu superyó. Por supuesto, dada la fuerza del superyó, podría ser racional seguir sus dictados en el sentido de que, de otro modo, te castigaría con intensos sentimientos de culpa y te haría infeliz. Incluso parecería que también deberías ser capaz de preguntar si sería o no mejor que abandonaras el superyó, por ejemplo, mediante un psicoanálisis intenso.

Pero no está claro que esta pregunta tenga sentido. Tus estándares judicativos se derivan de tu superyó; por lo tanto, a menos que haya un conflicto interno entre esos estándares, parecería que tienes que estar de acuerdo en que es deseable formar y conservar el superyó.

Lo que te hace sentir incómodo con respecto a esto es la sospecha de que en una sociedad diferente habrías internalizado diferentes estándares morales. En ese caso, una vez que has hecho público preguntar si era racional actuar en concordancia con *aquellos* estándares, tendrías que formular tu juicio sobre la base de aquellos mismos estándares y, nuevamente, a menos que hubiera un conflicto interno en los estándares mismos, parecería que deberías estar de acuerdo en que era deseable internalizar y conservar *aquellos* estándares.

Ahora bien, alguien que haya internalizado de este modo estándares morales que difieren de aquellos que tú has internalizado no puede decirse que actúa irracionalmente cuando actúa de acuerdo con sus estándares, aun cuando su acción sea incorrecta según los tuyos. Ya que, sobre la base de sus estándares, él juzgará que eres tú, y no él, quien actúa incorrectamente, y ¿quién dirá cuál de los dos está en lo correcto? En consecuencia, parece posible que cuando formas tu superyó internalizas principios que no están determinados únicamente por la racionalidad, dado que alguien igualmente racional, en otra sociedad, podría internalizar principios que difieren de los tuyos.

Esto indica una manera en la que podrías esperar alcanzar un juicio independiente de peso acerca de los principios morales que has internalizado. Podrías esperar descubrir la medida en que estos principios contienen principios universales que deberían ser aceptados por todo ser racional y la medida en que tu moralidad contiene principios que van más allá de lo que es requerido únicamente por la racionalidad.

Podría hacerse manifiesto que cuando aprendes la moralidad aprendes dos cosas. En primer lugar aprendes las costumbres de tu sociedad. En segundo lugar, aprendes ciertos principios generales de racionalidad entre los cuales está el principio según el cual, en igualdad de circustancias, debes adherirte a las costumbres del grupo local. En ese caso, las diferencias en las costumbres de un grupo a otro no serían un signo de diferencias en los principios morales básicos. Habría en cambio una moralidad universal consistente en principios de racionalidad práctica. En un sentido importante, el relativismo moral sería falso.

En el próximo capítulo discutiremos esta idea de que la moralidad puede ser identificada con la racionalidad práctica.

Por otra parte, podría ser que las constricciones de la racionalidad práctica fueran insuficientes para producir una moralidad universal. En ese caso, podrían ser verdaderas varias cosas. Una posibilidad es que aprender moralidad consiste en desarrollar principios morales

personales que aceptas como tus propios principios, no importa lo que los demás acepten como sus principios. En ese caso, aunque los principios que aceptas tienen un origen social en virtud de tu educación, ése es un hecho acerca de su origen más que acerca de su contenido. Tus principios serían tuyos en el sentido en el que tus gustos y disgustos son tuyos, aunque en alguna medida reflejen tu entorno social y tu educación. En el capítulo 7 discutiremos esta idea de que los principios morales son principios personales.

Otra posibilidad es que llegues a aceptar los principios morales básicos en tanto principios aceptados en un grupo y que diferentes grupos podrían aceptar diferentes principios. Esto representaría una forma social de relativismo moral. Entonces, la fuente social de la moralidad se reflejaría en el propio contenido de los principios morales. En los capítulos 8 y 9 discutiremos esta idea de que los principios morales son aceptados como principios que son aceptados por parte de los miembros de un grupo particular.

NOTA ACERCA DE LECTURAS ADICIONALES

La obligación es discutida por H.L.A. Hart, "Legal and Moral Obligation" en A.I. Melden, *Essays in Moral Philosophy*, University of Washington Press, Seattle, Wash., 1958.

Philippa Foot discute el principio del doble efecto en "Abortion and the Doctrine of Double Effect", en James Rachels (comp.), *Moral Problems*, Harper & Row, Nueva York, 1971.

El modo en que el "debería" moral presupone un sistema de leyes es descrito por G.E.M. Anscombe en "Modern Moral Philosophy", *Philosophy*, vol. 33, 1958.

El punto de vista de Freud respecto de la moralidad y el superyó se presenta en *El malestar en la cultura*, Alianza Editorial, Madrid, 1970.

VI

LEYES DE RAZÓN

1. *La fuente de la motivación moral*

En nuestro pensamiento moral cotidiano presumimos que la gente conoce ciertos principios morales y que puede algunas veces saber lo que debería hacer de acuerdo con esos principios. No pensamos que la gente sabe siempre lo que debería hacer moralmente, pero sí pensamos que algunas veces es posible que sepa tal cosa, y no censuraríamos moralmente a nadie por hacer algo que no podría haber sabido que era incorrecto. Así, por ejemplo, suponemos que no puede ser verdad que alguien debería moralmente haberte ayudado en tus momentos de necesidad si no había manera de haber sabido que necesitabas ayuda. De manera similar, normalmente suponemos que los principios morales sólo son obligatorios para las criaturas capaces de conocer esos principios. Ésa es una razón por la que suponemos que los animales y los débiles mentales no están sujetos a la moralidad; los suponemos incapaces de conocer los principios morales relevantes. Si resultara correcta alguna forma de nihilismo moral y el conocimiento moral resultara imposible, de modo que uno no pudiera saber nunca lo que debería hacer moralmente, se seguiría, de acuerdo con nuestros modos usuales de pensar, que nunca nadie podría hacer algo moralmente incorrecto. Aun la forma más moderada de nihilismo, la cual concede que hay hechos morales pero niega que podamos conocerlos (véase el capítulo 3), socavaría completamente, en consecuencia, la moralidad tal como la pensamos ordinariamente. Así, o bien el conocimiento moral es posible, o la moralidad tal como la concebimos ordinariamente es, en las palabras del filósofo alemán Immanuel Kant (1724–1804), "una noción vana y quimérica".

Pero tal como lo hemos visto en los capítulos 1 y 2, si el conocimiento moral existe es un tipo de conocimiento que no parece estar fundado en la observación de al menos un aspecto importante. Los hechos morales (si es que hay alguno) no parecen ayudar a explicar por qué alguien realiza las observaciones particulares que realiza. Esto indica que el conocimiento moral debe ser un tipo de conocimiento que puede adquirirse de un modo que no es la observación. (Este tipo de conocimiento se denomina a veces conocimiento *a priori*.) Una posibilidad sería que los principios morales básicos fueran autoevidentes, como los principios básicos en matemáticas. Es cierto, tal como lo advertimos en el capítulo 1, que los principios matemáticos cumplen un papel al ayudar a la gente a explicar por qué observa lo que observa, mientras que aparentemente los principios morales no lo hacen; pero quizá esta diferencia, si es real, no sea crucial.

Además, si hubiera tal *conocimiento a priori* de los principios morales, este conocimiento también representaría una fuente de *motivación a priori* que no depende del conocimiento basado en la observación. Ya que, tal como hemos aprendido del emotivismo, aceptar un principio moral es estar a favor o en contra de algo; y esto es estar motivado en algún grado para tratar de hacer algo o de evitar que algo ocurra.

Kant ofrece un argumento para una conclusión similar pero más fuerte aún. Dice que, de acuerdo con nuestra concepción usual de la moralidad, los principios morales básicos obligan a todos los seres racionales, de modo que, si existen los ángeles o los marcianos inteligentes, ellos también deben mantener sus promesas, no deben mentir, no deben causar sufrimiento innecesario, etc. Pero tal como ya hemos advertido, también es parte de nuestra concepción usual de la moralidad el hecho de que los principios morales obligan sólo a quienes conocen esos principios. Por lo tanto, si la sugerencia de Kant es correcta, se sigue de nuestra concepción usual de la moralidad que los principios básicos pueden ser conocidos por todos los seres racionales. Dado que aceptar un principio moral es estar motivado al menos en alguna medida a actuar en concordancia con tal principio, esto significa que nuestra concepción usual de la moralidad tiene que presuponer una fuente *a priori* de la motivación moral en la razón y en la racionalidad misma. Tiene que ser posible, para todo ser racional, estar motivado a actuar moralmente simplemente pensando y razonando, sin apelar al conocimiento adquirido a través de la observación. La conclusión de Kant es que, o bien la sola razón constituye tal fuente de

motivación, o la moralidad, tal como la entendemos ordinariamente, es una noción vana y quimérica.

Ahora bien, en la concepción humeana o aristotélica del razonamiento práctico que mencionamos antes (en el capítulo 3), el razonamiento práctico es siempre un razonamiento de medios a fines. Comienzas con uno o más fines y mediante el razonamiento determinas los mejores medios para obtener aquellos fines. Esto te lleva a adoptar aquellos medios como nuevos fines intermedios. Repites este proceso y, finalmente, de esta manera llegas a adoptar como tal medio, o como un fin intermedio, algo que está inmediatamente en tu poder, y que entonces haces. En esta óptica, el razonamiento práctico sirve para ayudar a satisfacer tus deseos existentes de antemano. El deseo puede ser una fuente última de motivación pero el razonamiento práctico no, dado que la función de ese razonamiento es simplemente la de organizar y canalizar la motivación existente. En esta óptica, entonces, hay sólo un tipo de irracionalidad práctica, a saber, no llegar a los medios que satisfarían tus fines. No puede haber irracionalidad en tus fines últimos en sí mismos. Todos los deseos son posibles, al menos en teoría. Ninguno es irracional. Hume dice, por ejemplo, que no es contrario a la razón preferir la destrucción del mundo al rasguño de tu dedo. Eso sería irracional únicamente si te preocuparas más por el mundo que por tu dedo, y puedes no preocuparte demasiado por el mundo.

Como hemos visto, Kant argumenta que o la razón es una fuente de motivación o la moralidad es una ilusión. Según la concepción del razonamiento práctico de Hume y de Aristóteles, la razón no puede ser una fuente de motivación. Por lo tanto, *o* la moralidad es una ilusión *o* la concepción del razonamiento práctico en Hume y en Aristóteles es inadecuado *o* hay algo incorrecto en el argumento de Kant. Revisando el argumento vemos que éste supone que tres cosas son esenciales a la moralidad tal cual la concebimos:

1. que un principio moral sólo obliga a quien puede conocer ese principio;
2. que los principios morales básicos obligan a todos los seres racionales; y
3. que aceptar un principio moral es estar motivado al menos en algún grado a actuar en concordancia con él.

Ahora bien, quizás la premisa más débil sea la 2, pues, como veremos más adelante: que los principios morales obliguen a todos los seres racionales no es obviamente *esencial* para nuestra concepción ordina-

ria de la moralidad. Por lo tanto, si deseamos mantener la concepción de Hume y Aristóteles del razonamiento práctico como razonamiento de medios a fines, podemos desarmar el argumento de Kant rechazando 2.

Habría aún un argumento para la concepción más débil según la cual la moralidad presupone conocimiento *a priori* de los principios morales conjuntamente con algún tipo de motivación *a priori*. Este argumento supone que 1 y 3 son esenciales a la moralidad tal como la concebimos:

1. que un principio moral obliga sólo a quien podría conocer ese principio;

y

3. que aceptar un principio moral es estar motivado al menos en algún grado a actuar en concordancia con él.

El argumento también supone:

4. que, si existe cualquier conocimiento moral, existe alguno que carece de sostén observacional.

Dados estos supuestos, se sigue que la moralidad presupone conocimiento *a priori* de los principios morales y una motivación *a priori* para actuar en concordancia con esos principios. Pero este argumento no está en conflicto con la concepción de Hume y Aristóteles del razonamiento práctico, porque la motivación moral puede tener una fuente *a priori* sin tener una fuente sólo en la razón. Por ejemplo, el conocimiento relevante y la motivación podrían estar incorporados en la gente como resultado de la evolución, de modo que la gente tendría un conocimiento innato de los principios morales, un conocimiento que estaría necesariamente vinculado a ciertas metas o deseos innatos. La idea de Hume de que la gente nace con una tendencia innata a la simpatía hacia los demás puede ser suficiente para esto. Y, tal como lo veremos, también hay otras posibilidades.

2. *¿Por qué preocuparse por otro?*

Por lo tanto, ni el argumento de Kant ni éste nos *fuerzan* a elegir entre la moralidad y la concepción de Hume y Aristóteles del razonamiento práctico. Ninguno de los dos argumentos *prueba* que la motivación debe tener su fuente en la razón misma si es que la moral no es una ilusión. Por otro lado, por lo que hemos dicho hasta aquí, puede ser irracional no preocuparse por la gente, y no sólo por razones de interés

propio. Sea o no *provechoso* a la larga desarrollar un interés genuino por los demás, es posible que tengas una razón para preocuparte por ellos que no surja de consideraciones de interés propio. Puede ser tan irracional no preocuparte por otra gente como no preocuparte por ti mismo. Quizá las razones para preocuparse puedan encontrarse en la misma naturaleza de la razón y de la moralidad. Ésta, en alguna medida, es la sugerencia de Kant, que ahora tenemos que examinar.

Consideremos un caso específico. Austin P. Jones, un individuo completamente interesado en sí mismo, está paseando placenteramente por una calle poco concurrida y aislada. Adelante ve un automóvil parado y el conductor, quien aparentemente no tiene gasolina, trata de detenerlo. Nos preguntamos si Jones tiene alguna razón para ayudar al conductor detenido. ¿Sería irracional de su parte no hacer caso al conductor y continuar su camino?

Ahora bien, puede ser verdad o no que Jones hubiera tratado de desarrollar tempranamente en su vida un interés por los demás. Quizá su vida sería más feliz ahora si lo hubiera hecho. Y en ese sentido, quizá, es irracional de su parte no preocuparse por el conductor detenido. Pero ése no es el punto. Debemos suponer que ahora es demasiado tarde para que Jones cambie sus "colores". Sus maneras están fijas y, tal como es ahora, no tiene interés por los demás. No es que le disgusten particularmente; sólo que no se preocupa por ellos de una manera u otra. Por lo tanto, en la instancia presente, observa el pedido del conductor detenido con desaprensiva curiosidad y sin sentir la más mínima motivación de ayudarlo. La pregunta que debemos responder es si, dada la falta de preocupación de Jones hacia los demás, él tiene no obstante una razón para detenerse y ayudar. ¿Le dan las necesidades del conductor detenido una razón de ese tipo, aunque no se preocupe por el conductor detenido? ¿Es irracional de su parte pasar de largo sin detenerse?

Podríamos intentar argumentar así con Jones: "Si la situación se invirtiera serías tú quien necesitaría ayuda y tus necesidades le darían a él una razón para ayudarte. Por lo tanto, por paridad de razonamiento, sus necesidades deben darte una razón para ayudarlo." Ése podría ser un buen argumento, si Jones concordase en que su propia necesidad de ayuda daría a otra persona una razón para ayudarlo. Pero Jones puede no estar de acuerdo. Puede sostener, por ejemplo, que sólo el interés propio puede dar a una persona una razón para hacer algo.

Por otra parte, algunas veces convencemos a una persona interesada en sí misma de que haga algo mediante un argumento en el que le

pedimos que considere qué ocurriría si la situación fuera a la inversa. Esto puede acaecer aunque sepa muy bien que la situación nunca será la contraria, al menos respecto a quien necesita ayuda en esta ocasión. Por lo tanto, cuando usamos tal argumento para convencer a una persona interesada en sí misma, no necesitamos convencerla de que en esta ocasión es de su interés prestar ayuda. En lugar de ello, la convencemos de que tiene una razón para ayudar, una razón que no se basa en el autointerés. Es sorprendente que, razonando de este modo, podamos hacer que una persona autointeresada actúe por tal razón.

Esto conduce a dos conclusiones. En primer lugar, que el hecho de que quisieras que otra persona te ayudara si la situación se invirtiera, podría ser en sí misma una razón para que lo ayudaras en esta situación, al margen de las consideraciones de autointerés. Si esto es así, Jones tiene una razón para detenerse y ayudar al conductor detenido, lo advierta o no, y actúa irracionalmente si no para. En segundo lugar, que el razonamiento puede hacerte actuar no sólo revelándote en qué medida tu acción es un medio para lo que ya deseabas, sino también mostrándote que tienes una razón para actuar, lo quieras o no lo quieras.

3. *Los deseos en tanto datos*

En otras palabras, esto sugiere una teoría del razonamiento semejante a la de Kant o semejante a una teoría que fue defendida elegantemente por el filósofo norteamericano contemporáneo Thomas Nagel, en *The Possibility of Altruism*. En tal teoría, la razón es la fuente de motivación. Al decidir qué hacer, te preguntas qué razones tienes para hacerlo. Algunas de estas razones pueden derivar de tus deseos actuales, pero otras no. Entonces tratas de hacer aquello para lo que tienes las mejores razones.

Nagel ofrece como ejemplo el hecho de que puedes comprarte comestibles aunque no tengas hambre cuando lo haces. Compras los comestibles porque sabes que posteriormente vas a tener hambre. Supones que tu deseo futuro de comida te da una razón para comprar ahora. De acuerdo con la concepción de Hume y de Aristóteles esto sería imposible. Para este punto de vista, un deseo futuro no puede darte una razón para hacer algo. Pero en la concepción kantiana de Nagel referida al razonamiento práctico, tú puedes tener una razón para hacer algo que no se deriva de ningún deseo antecedente que tengas concurrentemente. En realidad, según Nagel, es un principio de la

racionalidad práctica que, si vas a tener cierto deseo, tienes una razón para hacer algo que facilitará la satisfacción de ese deseo, tengas o no tengas un deseo actual.

Hume o Aristóteles tendrían que decir que estarás motivado a comprar comestibles sólo cuando tengas algún tipo de deseo precedente, tal como el deseo de que se satisfagan tus deseos futuros. Pero, como advierte Nagel, esto amenaza con volver circular la concepción de Hume y Aristóteles. Pues, ¿qué criterio ha de usarse para determinar si existe tal deseo precedente? Si la prueba es simplemente si hay alguna motivación, entonces es trivialmente verdadero que cuando hay una motivación hay un deseo precedente; pero ésta no es un objeción a la teoría kantiana de Nagel, quien advierte que este deseo de satisfacer tus deseos futuros, si es correcto decir que existe un deseo tal, no es un deseo del mismo tipo que el deseo de comida o de agua. Nagel argumenta que su fuente no consiste en ningún estado transitorio del cuerpo, sino más bien en la naturaleza de la identidad personal y de la razón misma. De acuerdo con Nagel es un deseo que sería irracional no poseer; eres irracional si no te preocupas por el futuro, sabiendo que en el futuro vas a preocuparte.

La concepción de Hume y Aristóteles del razonamiento práctico y la concepción kantiana de Nagel tratan los deseos básicos de una manera diferente. Para Hume y Aristóteles, deseos básicos como el hambre y la sed son fuerzas o energía que la razón dirige y organiza. En su concepción, la razón en sí misma no provee energía ni fuerza. En la teoría más kantiana de Nagel, los deseos básicos no son fuerzas o energía, sino un tipo de datos. Cuando razonas, tu información incluye el hecho de que tienes o tendrás ciertos deseos. Estos deseos influyen en lo que querrás hacer porque los tomas para que te den razones para hacer cosas y tú tratas de hacer aquello para lo cual tienes razones. Pero los deseos actuales no te dan más razones para hacer algo que los deseos futuros y es irracional satisfacer únicamente los deseos presentes, a expensas de los deseos futuros.

Ahora bien, de acuerdo con Nagel, los deseos de otra gente pueden darte una razón actual para hacer algo, tal como lo hacen tus propios deseos futuros. En la concepción de Nagel, tienes una razón para hacer todo lo que promueva la satisfacción de algún deseo, no importa que sea tuyo o de otra persona. Si puedes ayudar a una persona y no tienes ninguna razón para dejar de hacerlo, entonces es irracional de tu parte no ayudar a esa persona, así como lo es que no tomes precauciones para la satisfacción de tus propios deseos futuros. En

consecuencia, en la concepción de Nagel, Jones tiene una razón para detenerse y ayudar al conductor detenido, y sería irracional de su parte seguir conduciendo sin ayudarlo. Hume y Aristóteles no estarían de acuerdo; dirían que los deseos de otra gente pueden darte razones para hacer cosas sólo si tienes un deseo previo: por ejemplo, que se satisfagan los deseos de otra gente. Pero Nagel sostiene que este tipo de deseo previo no es un deseo básico del cuerpo, tal como el hambre y la sed, sino que es, más bien, una reflexión acerca del modo en el que funciona el razonamiento práctico; por tanto, de acuerdo con Nagel, es irracional no tener ese deseo. Si Nagel está en lo cierto, un interés por los otros está incorporado en la estructura misma del razonamiento práctico, y, en la medida en que la motivación moral refleja tal interés por los demás, la fuente de dicha motivación debe encontrarse en la razón misma.

Pero deben indicarse dos cosas acerca de la teoría de Nagel. Primero, no ha probado que sea irracional no preocuparse por la demás gente o que sea irracional no preocuparse acerca de lo que te ocurriría en el futuro. Segundo, la concepción de Nagel de la fuente de la motivación moral como derivada de una preocupación racional por los intereses de los demás no explica por sí misma por qué debería de haber algo semejante a la ley moral. Por ejemplo, por qué deberíamos distinguir, tal como lo hacemos, entre el médico que despedaza al paciente sano para salvar a los otros cinco pacientes y el doctor que salva a cinco pacientes y no a un paciente (capítulo 1) o el conductor del tranvía que mata a un trabajador en un riel lateral antes que a los cinco trabajadores del riel principal (capítulo 5). En la medida en que los intereses de otra gente suministren a un agente razones para promover aquellos intereses, es difícil ver sobre qué bases podrían distinguirse estos casos.

4. *La prueba de Kant*

Kant está peculiarmente interesado en poner de manifiesto el modo en que la moralidad involucra la ley moral, aunque nunca (hasta donde sé) se dedica a la cuestión recién planteada. La idea básica de Kant es que actuar moralmente es actuar respetando principios que todo ser racional aceptaría. Estos principios, según Kant, componen la ley moral. No basta con actuar siguiendo esos principios; no es suficiente "hacer lo correcto". El dueño de un almacén que regularmente da el vuelto correcto aun a los niños, porque hacerlo es bueno para el

negocio, no necesariamente está actuando moralmente, aunque está actuando como si lo estuviera haciendo. Que un acto sea o no moral depende de por qué lo haces. Tienes que hacerlo porque es correcto hacerlo. Si lo haces sólo por otras razones, tu acto no tiene mérito moral. Kant dice que un acto tiene mérito moral sólo si se hace porque es correcto, porque es exigido por la ley moral; es decir, porque es exigido por una ley moral que todos los seres racionales deben aceptar.

Kant supone que toda vez que actúas por una razón, tu acción depende siempre de un principio básico o último. (Frecuentemente denomina "máxima" a este principio.) Por ejemplo, si actúas a modo de satisfacer tus deseos, tu principio básico o máxima es actuar a modo de satisfacer tus deseos. Ahora bien, según Kant, si vas a actuar moralmente, tu principio básico o máxima debe ser un principio que aceptas en tanto principio para todos los seres racionales. En la terminología de Kant, debes "querer" tu principio como una ley universal para todos los seres racionales. En consecuencia, Kant supone que una prueba del principio conforme al cual actúas, es si estás de acuerdo en que todos actúen según ese mismo principio y, para determinar si estarías de acuerdo en eso, siempre deberías actuar como actuarías si la "máxima" o principio último según el cual actúas se convirtiera, por milagro, en una ley universal conforme a la cual todos actuaran siempre.

Actuar de tal manera, según Kant, es actuar a partir de la ley moral. Por lo tanto, también estás actuando como si todos, por milagro, actuaran a partir de la ley moral. Estás actuando, en consecuencia, como lo harías en un mundo en el que todos actuaran, en todo momento, según la ley moral. En tal mundo, cada persona tendría un mérito moral incondicional, de acuerdo con Kant, quien argumenta que lo único que tiene mérito moral incondicional es que alguien trate de hacer lo correcto. Algo de mérito moral incondicional es un "fin" en la terminología de Kant. Ésta es una de las razones por las que Kant dice que, al actuar moralmente, actúas como si estuvieras legislando para un Reino de Fines. Esto equivale a algo semejante a la Regla de Oro "Haz a los demás lo que quisieras que te hicieran a ti", excepto por que la regla de Kant concierne únicamente a los principios últimos de la acción: los principios últimos según los cuales actúas deberían ser principios que tú elegieras tener como los principios últimos según los cuales todos actuaran.

Tal como lo entiende Kant, esto involucra un tipo de regreso o de círculo, ya que el principio que elegirías para que todos actúen es también el principio que está detrás de la elección misma. En otras palabras, en la concepción de Kant tú actúas racionalmente y en consecuencia moralmente al hacer *H* sólo si hay un principio *P* tal que:

1. haces *H*;
2. tu principio último al hacer lo que haces en 1 es *P*;
3. elegirías *P* como un principio último según el cual todos deberían actuar, y
4. tu principio último al hacer lo que haces en 3 es *P*.

Si eliges que ciertos principios sean adoptados por todos porque supones que sería mejor para tus intereses que se adoptaran dichos principios, entonces tu principio último es en realidad puro autointerés, y por lo tanto, el problema es si elegirías por puro autointerés que todos actuaran siempre por puro autointerés. Presumiblemente no elegirías eso, dado que no está en tu interés que todos actúen siempre según su propio interés. Por otra parte, podrías elegir un principio semejante a la Regla de Oro como principio último, pues ello podría proporcionarte una razón para elegir que todos actúen según ese principio. Si eliges la Regla de Oro, tu razón para elegirla tendría que ser que elegirla para que todos la sigan es hacer a los otros lo que quisieras que te hicieran a ti. Tus razones últimas para elegir que ciertas razones sean las razones últimas de todos deben ser, en la concepción de Kant, esas mismas razones.

Tus principios últimos tendrían que *incluir* el interés propio, ya que (Kant piensa) todo ser racional es necesariamente autointeresado. Pero tu principio tiene que ir también más allá del mero autointerés, dado que no es de tu propio interés que todos actúen únicamente según sus propios intereses. En consecuencia, tu principio tiene que dar cabida adecuada al altruismo y a otra motivación no autointeresada de modo que, dados estos motivos adicionales al autointerés, fuera racional elegir que todos siempre actuaran según tales motivos. Kant piensa que hay un único resultado de esta prueba y que el principio resultante incluirá no sólo el interés propio y un altruismo limitado, sino, además, reglas acerca de la propiedad, el prometer, etc.; de hecho, los principios usuales de la moralidad. Pero esto no demuestra que su prueba produzca un único resultado, de la manera indicada.

En realidad, por razones que no es necesario considerar aquí, Kant cree que nunca se puede *saber* que hay una ley moral, una ley que

sólo se deriva de la razón. Sin embargo, argumenta que toda vez que actúes tienes que suponer implícitamente que tal ley existe. Porque toda vez que estás decidiendo qué hacer tienes que pensar en tu decisión inminente a la manera kantiana antes que a la manera humeana o aristotélica. Tienes que ser capaz de suponer que eres libre de decidir de una u otra manera. Tienes que suponer que vas a tomar tu decisión por ciertas razones. Por lo tanto, no puedes suponer que tu decisión es simplemente el resultado de un número de fuerzas, representadas por tus deseos. Pensar tus deseos de este modo, simplemente como fuerzas (tal como en la concepción de Hume y de Aristóteles), sería pensarlos como compulsiones que no son controladas por ti. Es cierto que algunas veces vemos nuestros deseos como compulsiones; pero normalmente no lo hacemos. Normalmente, tratas tus deseos como datos de la manera descrita por Kant y Nagel. Te piensas a ti mismo como actuando por una razón. Si no lo hicieras, no podrías suponer que estabas decidiendo algo. Además tienes que suponer que estás actuando racionalmente, que tus razones últimas son buenas razones.

Kant no piensa que puedas saber si tú o alguien alguna vez decide actuar en este sentido. Piensa que es posible que tus acciones sean siempre el resultado ciego de deseos que compiten y que nunca actúes por una u otra razón. No obstante, argumenta que normalmente supones que hay razones por las que haces lo que haces. Tu suposición según la cual actúas por razones es lo que Kant denomina un "postulado práctico". No es susceptible de prueba pero es necesario para la concepción de ti mismo en tanto agente.

Ahora bien, Kant argumenta finalmente que esto implica que cuando actúas, tienes que postular la ley moral como la piedra de toque de la racionalidad. Ya que, cuando actúas, tienes que suponer que estás actuando por una razón y, en consecuencia, tienes que suponer que hay un principio último o una razón detrás de tu acción. Más aún, tienes también que suponer que es racional de tu parte actuar según tu principio último o razón, lo cual significa, según Kant, que tienes que suponer que tu razón última es una razón para todo ser racional. (Porque si tu principio consiste en hacer *H*, y piensas que existe un ser racional para quien hacer *H* no es racional, tu principio último no es simplemente hacer *H*, sino hacer *H* si eres un ser de cierto tipo.) Por lo tanto, tienes que estar dispuesto a aceptar no sólo tu propio actuar según esta razón última, sino también el de todos los demás. En consecuencia, concluye Kant, cuando actúas tienes que querer tu

principio último o razón (o "máxima") en tanto ley universal para todos los seres racionales; de otra manera, eres irracional. Pero, advierte Kant, ésta es la ley moral según él la entiende; así concluye que, cuando actúas, debes postular la ley moral como una prueba última de la racionalidad.

5. *Objeciones a Kant*

El argumento de Kant depende de una ambigüedad que concierne a lo que está involucrado en la aceptación de un principio. Supongamos que tu principio último es tratar de maximizar tu propia felicidad. Entonces tienes que pensar que este principio es racional; con otras palabras, tienes que pensar que es racional de tu parte tratar de maximizar tu propia felicidad. Dado que éste es tu principio último, tienes que pensar también que sería racional que todos trataran de maximizar su propia felicidad. (Ya que si pensaras que sólo para la gente del tipo *Q* sería racional tratar de maximizar su propia felicidad, tu principio último no sería simplemente tratar de maximizar tu propia felicidad sino más bien, por ejemplo, tratar de maximizar tu propia felicidad si eres una persona del tipo *Q*.) En consecuencia, hay un sentido en el que tienes que aceptar que todos actúen según tu principio básico; pues tienes que concordar con que es racional de su parte hacerlo. Pero no se sigue que tengas que aceptarlo en el sentido de que tienes que querer que todos actúen según ese principio; por lo tanto, Kant no ha mostrado que haya un error en la idea de que sería racional para todos tratar de maximizar su propia felicidad, aunque tú no quieras que lo hagan. En consecuencia, Kant no ha mostrado que haya nada irracional en actuar puramente según el interés propio.

La prueba ofrecida por Kant para los principios morales es igualmente defectuosa, en primer lugar porque, dado que puedes estar de acuerdo en que sería racional que todos actuaran según tu principio básico, sin que estés comprometido a querer que lo hagan, puedes también estar de acuerdo con esto sin tener que concordar en que sería racional, sobre la base de tu principio, lograr que todos actuaran según ese principio (si por milagro pudiéramos realizar tal cosa). Además, aun dejando a un lado el problema de si los principios morales correctos están limitados a aquellos principios que se aplicarían en un Reino de Fines, en el que todos actuaran moralmente y nunca nadie hiciera nada incorrecto. En el mundo real no todos actúan moralmente y los principios morales tienen que dar cabida a tal cosa. El pacifismo

total podría ser un buen principio si todos lo siguieran. Pero no todos lo hacen y, por lo tanto, no lo es.

Kant trata sin éxito de mostrar que la ley moral, en tanto ley de la razón, puede ser derivada de un análisis de qué es actuar libremente o qué es decidir hacer algo. Trata de mostrar que cuando te piensas en tanto agente tienes que postular principios de racionalidad que constituyen la ley moral. Es verdad que la concepción kantiana del razonamiento práctico (especialmente tal como fue elaborada y refinada por Nagel) es una concepción más plausible que la de Hume o la de Aristóteles, que trata a los deseos como fuerzas. Sin embargo, esta concepción no produce los resultados que Kant pretende.

Esto nos lleva al primer argumento de Kant, según el cual la motivación moral deriva sólo de la razón, o bien, la moralidad es una noción vana y quimérica. Pero, tal como lo advertimos antes, también hay una salida para ese argumento. Podemos negar simplemente la suposición de Kant de que los principios morales obligan a todos los seres racionales. Esto indica una concepción más relativista de la moralidad, en la cual los principios morales de cada persona son propios y los seres racionales no suponen necesariamente que comparten los mismos principios fundamentales. Examinaremos tal concepción de la moralidad en el próximo capítulo.

NOTA ACERCA DE LECTURAS ADICIONALES

La mejor exposición de la teoría moral de Immanuel Kant está en su *Fundamentación de la metafísica de las costumbres*. Hay una versión más legible y menos rigurosa en sus *Conferencias sobre ética*, basadas en apuntes de estudiantes. También son relevantes su *Crítica de la razón práctica* y *La religión dentro de los límites de la mera razón* (varias ediciones).

El comentario en inglés más útil acerca de la teoría moral de Kant es la defensa de Nagel de una concepción similar en *The Possibility of Altruism*, Oxford University Press, Oxford, 1970.

VII

PRINCIPIOS PERSONALES

1. *Una ley moral personal*

Muchos filósofos han sido muy influidos por la explicación de la moralidad ofrecida por Kant sin estar de acuerdo con su concepción acerca de los poderes de la razón. Aunque dichos filósofos no creen que los principios que deben ser aceptados por todos los seres racionales sean lo suficientemente fuertes como para constituir lo que normalmente concebimos como la moralidad, sí piensan que el análisis de Kant es en gran medida correcto en otros aspectos. Por ejemplo, concuerdan en que cuando actúas, tienes que pensar que lo haces según razones y no puedes pensar que simplemente eres arrastrado por tus deseos más intensos, pues de otro modo tendrías que pensar que estás actuando por compulsión o incluso que no estás actuando en absoluto. También concuerdan en que actuar según razones es actuar según principios y que esos principios pueden proporcionarte razones sólo si son principios racionales.

Ahora bien, Kant pensaba, tal como hemos visto, que esto significaba que los principios relevantes habían de ser principios que todo agente racional aceptara; en consecuencia, sería irracional para cualquier agente no aceptar esos principios. Los filósofos poskantianos de los que estoy hablando no están de acuerdo con esto. Creen que las únicas constricciones racionales sobre un sistema de principios son los requerimientos de simple consistencia. Así, sostienen que diferentes personas pueden tener diferentes principios, principios que les dan razones para hacer cosas diferentes que están en conflicto, sin que ninguno de ellos sea irracional, en la medida en que los principios de cada persona sean internamente consistentes.

Supongamos que estos filósofos están en lo cierto acerca de los poderes de la razón. Entonces, si siguiendo a Kant limitamos los principios morales a aquellos principios que son requeridos por la razón, la moralidad debe reducirse a la mera consistencia. Y, en realidad, ésta es la concepción del existencialista francés contemporáneo J.P. Sartre, quien piensa por esta razón que la moralidad equivale a lo que él denomina "buena fe"; actuar moralmente consiste únicamente en no ser hipócrita.[1] Por otra parte, podemos identificar la moralidad de cada persona con el sistema total de principios de esa persona y no sólo con aquellos principios requeridos por la razón. Esto es aproximadamente lo que hace el filósofo inglés contemporáneo que estudia el lenguaje, R.M. Hare, aunque, como veremos, él no considera ningún principio, cualquiera que sea, como un principio moral.[2] En su concepción, diferentes personas pueden tener diferentes moralidades, ya que pueden tener diferentes principios que aceptan como principios morales, sin ser irracionales.

Comparada con nuestra concepción usual de la moralidad, la concepción de Sartre parece demasiado débil y la de Hare demasiado fuerte. La de Sartre parece demasiado débil porque ordinariamente suponemos que en la moralidad hay algo más que la buena fe y el no ser hipócrita. Suponemos corrientemente que la moralidad requiere que no robes, que no mientas, que ayudes a quienes están en la desgracia, etc. Tales requerimientos tienen que involucrar algo más que ser simplemente auténtico, a menos que Kant esté en lo cierto respecto a los poderes de la razón.

Por otra parte, el punto de vista de Hare da contenido a la moralidad, pero no de la manera que ordinariamente pensaríamos correcta. De acuerdo con nuestro punto de vista corriente, la moralidad misma impone requerimientos a una persona; y el hecho de que alguien decida adoptar cierto principio como suyo no lo convierte en uno de sus principios morales, a menos que lo adopte porque piense que la moralidad requiere esa acción que concuerda con tal principio. Además, los que Hare considera como los principios morales de una persona, no precisan coincidir con lo que ordinariamente tomaríamos como los principios de la moralidad. No hay razón para adoptar principios contra la mentira o el robo y a favor de la ayuda a quienes están en

[1] "Existentialism is a Humanism", en Kaufmann (comp.) *Existentialism: From Dostoevsky to Sartre*, Meridian, Nueva York, 1956.

[2] *The Language of Morals*, The Clarendon Press, Oxford, 1952.

la miseria. Según el punto de vista de Hare, los principios morales de una persona pueden ser muy diferentes de lo que ordinariamente pensaríamos como los principios de la moralidad.

Pero los conflictos con nuestra concepción de la moralidad no pueden refutar la teoría de Hare ni la de Sartre, ya que ambos filósofos tienen la opción de contestar que nuestra concepción ordinaria de la moralidad presupone un punto de vista falso de la razón. Si los requerimientos de la razón son menores de lo que ordinariamente suponemos, tal como lo creen Hare y Sartre, entonces es posible que nuestra concepción ordinaria de la moralidad no pueda ser mantenida. Quizá la moralidad, tal como la concebimos ordinariamente, sea una "noción vana y quimérica", una ilusión. Sin embargo, esto en ningún sentido significaría que tendríamos que prescindir de ella. Y puestos a elegir, que decidiéramos hablar a la manera de Sartre o a la manera de Hare podría tornarse en una cuestión puramente verbal.

2. *La teoría de R.M. Hare*

La teoría de Hare es más elaborada que la de Sartre; considerémosla, pues, con más detalle. Hare postula que los principios morales básicos de una persona pueden ser expresados como imperativos generales de la forma "Para cualquier persona *P*, en todo momento en que *P* esté en las circunstancias *C*, ¡que *P* haga *H*!" Por ejemplo: "¡que toda persona que ha prometido hacer algo, haga lo que ha prometido hacer!" Se supone que estos principios son generales y que no tienen excepciones. Se supone que son generales porque están destinados a todos. Se supone que no tienen excepciones pues deben sostenerse en toda situación que pudiera surgir, sin excepción, incluyendo las situaciones meramente hipotéticas. En consecuencia, todas las excepciones potenciales deben estar "incorporadas" con anterioridad a los principios. Por ejemplo, si incorporamos excepciones potenciales a nuestro principio referido acerca de cumplir las promesas, podríamos terminar con algo semejante a esto:

> ¡Que toda persona que ha prometido hacer algo, si su promesa no fue obtenida con dolo o mediante coerción, haga lo que ha prometido hacer, a menos que pueda hacer algo más que sea incompatible con el cumplimiento de su promesa y que ello preste un servicio significativo a alguien que necesite ayuda, en cuyo caso, que haga esa otra cosa!

En efecto, aun un principio tan elaborado es irremediablemente supersimplificado si se compara con los principios que la mayoría de las personas deberían tener que aceptar si su moralidad se basara realmente en principios de este tipo. (Volveremos a este punto en el capítulo 10.)

Ahora bien, de acuerdo con Hare, aceptar un principio general en tanto principio moral es (a) tener la intención de adherirse a él, y (b) tener la intención de tratar de lograr que los demás lo acepten. La condición (b) expresa la generalidad o universalidad del principio. No es sólo un principio para ti. No dice solamente, por ejemplo: "Que en todo momento en que esté en las circunstancias *C*, ¡haga yo *H*!"

Tal como observa Hare, esto constriñe tus principios morales. Éstos no pueden dirigirse simplemente a la promoción de tu propio y estrecho interés, dado que tienes que estar a favor de que tus principios sean seguidos no sólo por ti, sino también por todos los demás —y no sólo en la situación presente sino también en las situaciones meramente hipotéticas en las que se revierten tus posiciones y las de los demás. No obstante, no hay un único conjunto racional de principios que satisfaga esta condición. Diferentes personas pueden aceptar conjuntos de principios diferentes y en conflicto que satisfagan (a) y (b).

Alguien podría aceptar solamente principios más débiles que satisfagan (a) pero que no satisfagan (b). Tendría la intención de seguir sus principios pero sin estar interesado en tratar de que los demás los siguieran. Por ejemplo, de esta manera podría aceptar principios netamente egoístas que no quiere que los demás sigan, pues no favorece su interés que los demás sigan principios egoístas. Tal persona no sería obviamente irracional. La razón no parece requerir que adoptes principios generales que satisfagan tanto (a) como (b). Es verdad que si piensas que actúas según razones (y no sólo como alguien que es una víctima de sus deseos), entonces tienes que suponer que las razones por las que actúas son buenas razones. Esto implica que tienes que suponer que quienquiera que estuviera en una posición idéntica tendría las mismas buenas razones para actuar como tú. Pero eso no implica que estés o debas estar dispuesto a hacer que otras personas actúen de manera similar por razones similares. De hecho, si eres egoísta, deberías tratar de impedir que los demás actuaran sobre la base de los que, según piensas, son los principios más racionales sobre los que pueden actuar. La racionalidad, por sí misma, no requiere que actúes a partir de principios que satisfagan la condición (b).

Pero si tus principios satisfacen sólo la condición (a), y no tienes la intención de tratar de lograr que otros los adopten, Hare no los considerará "principios morales". De acuerdo con Hare, tener principios morales es tener principios que pretendes seguir y que, también, pretendes hacer que otros sigan. Así, la racionalidad por sí misma no te obliga a tener principios morales en el sentido que Hare da a "principio moral".

Dada esta versión de los principios morales, Hare está listo para proporcionar una definición aproximada del "debería" moral. Dice que "*P* debería hacer *H*" significa algo como "el que *P* haga *H* va de acuerdo con los principios morales generales que por la presente suscribo", y esto a su vez significa algo como "los hechos, junto con los principios morales generales que por la presente suscribo, implican este imperativo dirigido a *P*: '¡Haz *H*!' ".

La frase "que por la presente suscribo" se necesita para permitir el desacuerdo moral. Pues supongamos que el análisis dice simplemente: "*P* debería hacer *H*" significa "el que *P* haga *H* va de acuerdo con los principios morales generales que suscribo. Entonces, si yo digo que *P* debería hacer *H* y tú dices que *P* no debería hacer *H*, según este análisis, ni siquiera estaríamos en desacuerdo. Pues estaría diciendo que hacer *H* por parte de *P* concuerda con los principios que yo suscribo, y tú estarías diciendo que no hacer *H* por parte de *P* concuerda con los principios que tú suscribes. Y estas observaciones no se contradicen entre sí, dado que tú y yo podemos tener diferentes principios.

Entonces, ¿cómo es que estamos en desacuerdo? Según Hare, estamos en desacuerdo en la actitud, pues estamos defendiendo principios que están en conflicto. Yo estoy defendiendo principios que requerirían que *P* hiciera *H*, mientras que tú estás defendiendo principios que requerirían que *P* no hiciera *H*. Cuando digo que hacer *H* por parte de *P* concuerda con los principios que suscribo por la presente, mi "por la presente" es un signo de que no sólo estoy describiendo mis principios sino que en la emisión misma estoy suscribiéndolos; en otras palabras, estoy abogando por ellos. Tú y yo diferimos en el hecho de que abogamos por cosas incompatibles.

Ahora bien, también diferimos si digo que *P* debe hacer *H* y tú simplemente niegas esto, sin decir que *P* debería no hacer *H*. En ese caso, nuestro desacuerdo no consiste sólo en que yo estoy abogando por principios que requieren que *P* haga *H*, mientras que tú estás abogando por principios que no requieren que *P* haga *H*, pues ello

podría significar que yo sólo estaba defendiendo principios más fuertes que los que estabas defendiendo tú, y mis principios pueden ser más fuertes que los tuyos sin estar en conflicto con ellos. No obstante, de hecho, cuando niegas que *P* debería hacer *H*, no sólo abogas por principios que no requieren que *P* haga *H*, sino que también abogas porque la gente no acepte como principios morales aquellos que tomados en conjunción requerirían que *P* hiciera *H*. Es decir, abogas porque la gente adopte, como conjunto total de principios, principios que no requieren que *P* haga *H*.

Por lo tanto, extendiendo el análisis de Hare a la negación de las oraciones morales que contienen "debería", vemos que "No es el caso que *P* debería hacer *H*" debería ser analizada como si significara algo semejante a "Yo suscribo por la presente y abogo porque la gente adopte, en tanto conjunto total de principios morales, principios *M* que son tales que no es el caso que hacer *H* por parte de *P* esté en concordancia con *M*." De manera similar, cuando el "debería" moral aparece incluido en una cláusula subordinada de una oración mayor, debería ser analizado de esta manera: "... *P* debería hacer *H*..." significa algo semejante a "Por la presente suscribo y defiendo que la gente adopte, como conjunto total de principios morales, principios *M* tales que... hacer *H* por parte de *P* esté en concordancia con *M*..."

Hare señala que hay otro uso del "debería" moral, en virtud del cual el hablante no suscribe los principios relevantes. Lo denomina el uso de comillas, dado que los principios son defendidos por otros. Por ejemplo: "Jones cree que *P* debería hacer *H*" no debe analizarse como "Por la presente suscribo y defiendo que la gente adopte, como conjunto total de principios morales, principios *M* tales que Jones crea que hacer *H* por parte de *P* está en concordancia con *M*". Se dice que Jones suscribe ciertos principios, no que lo hace el hablante. Para tal oración, el análisis debe ser semejante a "Jones suscribe, y se propone intentar que la gente adopte, un conjunto completo de principios morales, tal que él cree que hacer *H* por parte de *P* está en concordancia con *M*".

3. *La objeción existencialista a Hare*

Deseo considerar ahora una objeción al análisis de Hare desde el punto de vista de quien, como Sartre, restringiría los principios morales a principios de simple consistencia. Sartre y Hare coinciden en que puede ocurrir que, aunque tú y *P* sean perfectamente consistentes en

los principios que aceptan, tus principios impliquen que *P* tendría que hacer *H*, mientras que los suyos impliquen que no tendría que hacerlo. En tal caso, el análisis de Hare permitiría que dijeras de *P* que debe hacer *H*. La objeción a esto es que es raro decir tal cosa, dado que *P* no tiene razón para hacer *H*, al menos ninguna razón que derive de sus principios. Es parte de nuestro punto de vista ordinario que *P* debe hacer *H* sólo en los casos en los que *P* tiene razones para hacer *H*. Y no está claro de dónde podrían provenir esas razones excepto de los principios de *P*. Pues, ¿cómo pueden *tus* principios dar a *P* alguna razón para hacer *H* si *P* no comparte tus principios? Decir que *P* debería hacer *H* es decir, al menos, que *P* tiene alguna razón para hacer *H*. Pero el análisis de Hare permitiría decir que *P* debería hacer *H* aun cuando tuvieras que concordar en que *P* no tiene razón para hacer *H*. En consecuencia —continúa la objeción— el análisis de Hare tiene que ser inadecuado.

Por supuesto, podrías aceptar el principio (O) como uno de tus principios morales, según el cual sólo cuando se tiene una razón para hacer algo debería hacerse. En ese caso, si tus principios morales fueran consistentes no podrían implicar que *P* debería hacer *H*, aunque no tuviera una razón para hacer *H*. La dificultad para Hare es que no hay nada en su análisis que requiera que aceptes (O) como un principio moral, por lo que tiene que admitir la posibilidad de que no aceptes (O) y, en consecuencia, que tus principios morales a veces impliquen que *P* debe hacer *H* aunque no tenga una razón para hacer *H*. La objeción consiste en que esta posibilidad tendría que ser eliminada.

Es posible ofrecer dos réplicas a esta objeción. Una niega el principio (O) según el cual *P* debe hacer *H* sólo si *P* tiene una razón para hacer *H*. La otra dice, por el contrario, que en el sentido relevante *P* tiene una razón para hacer *H*. Consideremos esta réplicas sucesivamente.

Primero, ¿es verdad que decir que *P* debería hacer *H* implica que *P* tiene una razón para hacer *H*? Para responder esta pregunta debemos distinguir diversos sentidos de "debería". No existe tal implicación en lo que previamente hemos denominado el sentido evaluativo de "debería". Pero esto no es de ayuda para el análisis de Hare, dado que su intención es analizar el sentido moral de "debería", que no es el mismo que el sentido evaluativo. Cuando "debería" se entiende en su sentido moral, la implicación existe.

En el capítulo 5 señalé que hay al menos cuatro sentidos de "debería". En primer lugar, hay un "debería" de expectativa, como cuando

decimos que el tren ya debería de llegar pronto. Luego hay un "debería" de evaluación, como cuando decimos que debería haber más amor en el mundo, queriendo decir que sería algo bueno que hubiera más amor en el mundo. No significa que se espera que haya más amor en el mundo. Y al decir que el tren debería de llegar pronto, no significa necesariamente que sería bueno que el tren llegara pronto.

Estos dos sentidos de "debería", el "debería" de expectativa y el "debería" evaluativo, se usan para describir estados de cosas. "Debería" puede usarse también para caracterizar a un agente de una manera más específica. Podemos decir que un agente debería realizar cierto acto. Además, hay al menos dos sentidos diferentes en los que podemos decir esto. Podemos usar el "debería" moral para decir que una persona moralmente debería hacer algo; por ejemplo, al hablar de un ladrón que va a robar un banco, podríamos decir que debería abandonar su plan y volver a su casa. O podemos usar el "debería" de racionalidad; al hablar del ladrón, podríamos decir que debería entrar por la puerta trasera, dando a entender que sería racional de su parte hacer tal cosa para evitar a los guardias de la puerta delantera. Claramente, hay una diferencia en lo que "debería" significa cuando decimos que el ladrón debería irse a su casa y cuando decimos que debería usar la puerta trasera.

Ahora bien, quizá podemos estar de acuerdo en que hay una clara diferencia de sentido entre el "debería" de expectativa, el "debería" moral y el "debería" de racionalidad (aunque, si Kant estuviera en lo correcto, el "debería" moral coincidiría con el "debería" de racionalidad). Lo que es más difícil de ver es alguna diferencia entre el "debería" evaluativo y el "debería" moral. Pero consideremos el comentario "Rockefeller debería donar su dinero a los pobres". Esta aseveración es sutilmente ambigua entre las interpretaciones evaluativa y moral de "debería". Podría entenderse como un comentario acerca de lo que debería ser, esto es, que debería ser el caso que Rockefeller donara su dinero a los pobres. O bien la observación puede interpretarse como una observación acerca de Rockefeller en tanto agente, diciendo que existe un acto que moralmente él debería realizar: moralmente, Rockefeller debería donar su dinero a los pobres. En un caso, el hablante está diciendo que sería bueno cierto estado de cosas. En el otro caso, el hablante está diciendo que sería incorrecto por parte de Rockefeller no seguir cierto curso de acción. La segunda aseveración es más fuerte. Podrías formular la primera aseveración porque piensas que sería bueno que no existieran las diferencias en

cuanto a riqueza que existen actualmente. Podrías formular la segunda aseveración si piensas que Rockefeller obtuvo su dinero explotando a los pobres y que, en consecuencia, se lo debe. Si formulas la aseveración más fuerte usando el "debería" moral, implicas que Rockefeller hace algo moralmente incorrecto al no devolver el dinero. Si formulas la aseveración más débil, usando el "debería" evaluativo, no implicas necesariamente que Rockefeller esté equivocado, aunque implicas que cierta situación es incorrecta.

El "debería" evaluativo es usado para evaluar una situación: el que Rockefeller dé su dinero a los pobres. El "debería" moral es usado para formular un juicio acerca del tema de la oración, en este caso Rockefeller. En consecuencia, hay una manera de distinguir el "debería ser" evaluativo del "debería hacer" moral. Si el significado evaluativo es el correcto, la oración "Rockefeller debería donar su dinero a los pobres" es equivalente a "Debería ser el caso que Rockefeller donara su dinero a los pobres". Si el sentido moral es lo relevante, esta equivalencia no existe, pues que Rockefeller dé su dinero a los pobres es lo mismo que los pobres obtengan el dinero de Rockefeller, si lo que se tiene presente es el sentido evaluativo, la oración "Rockefeller debería donar su dinero a los pobres" es equivalente a la oración "Los pobres deberían obtener el dinero de Rockefeller". Si lo que se tiene presente es, por el contrario, el sentido moral, estas oraciones no pueden ser equivalentes, pues la primera dice lo que Rockefeller debería hacer mientras que la segunda dice lo que los pobres deberían hacer (si es que pueden).

Ahora bien, Hare se propone analizar "*P* debería hacer *H*", en la que "debería" está usado en su sentido moral, como opuesto a sus sentidos probabilístico, evaluativo o racional. Pero en el sentido moral de "debería", decir que *P* debería hacer *H* es implicar que *P* tiene razones para hacer *H*, ya que es inconsistente suponer que *P* debería (moralmente) hacer *H* cuando no tiene ninguna razón —ni siquiera una razón moral— para hacer *H*.

Este punto queda oscurecido si no se distingue claramente el "debería" moral del "debería" evaluativo. Al decir que Rockefeller debería donar su dinero a los pobres, podrías no estar diciendo ni siquiera en parte que Rockefeller tiene una razón para donar su dinero a los pobres. Podrías decir esto y concordar consistentemente en que Rockefeller no tiene una razón para donar su dinero a los pobres, si estás usando el "debería" evaluativo. Si sólo estás diciendo que debería ser

el caso que Rockefeller donara su dinero a los pobres, no estás necesariamente diciendo que existe alguna obligación moral por parte de Rockefeller para que done su dinero a los pobres y no estás implicando necesariamente que existan razones para que lo haga. No es inconsistente suponer que aunque Rockefeller no tenga una razón —ni siquiera una razón moral— para donar su dinero a los pobres, aún debería ser el caso que donara su dinero a los pobres: el mundo sería un lugar mejor. Pero si quieres decir que moralmente Rockefeller debería dar su dinero a los pobres porque, por ejemplo, obtuvo su dinero explotándolos, sí implicas que Rockefeller tiene una razón para donar su dinero a los pobres, a saber, que amasó su dinero explotándolos. El "debería" evaluativo, que es un "debería ser", no conlleva implicaciones acerca de tener razones para actuar por parte del agente. El "debería" moral, que es un "debería hacer", conlleva la implicación de que la persona que debe hacer algo tiene una razón para hacerlo.

4. *Las razones y los principios básicos*

Hemos descubierto que la primera manera de responder, en términos de Hare, a la objeción sartreana no funciona. Pero vemos que la segunda manera de responder podría funcionar. La objeción, como recordarás, es que Hare no puede dar cuenta de los juicios morales que podrías formular acerca de una persona cuyos principios difirieran de los tuyos. Decir que tal persona debería hacer algo es decir que tiene una razón para hacerlo. Pero Hare permitiría decir que una persona debería hacer algo aunque la persona en cuestión no tuviera una razón para hacerlo y aunque tuviera todas las razones para no hacerlo. La primera contestación a esta objeción sartreana intentó negar el principio (O), según el cual decir que alguien debería moralmente hacer algo, es decir que tiene una razón para hacerlo. He argumentado que esta respuesta parece plausible sólo si no se distingue el "debería hacer" moral del "debería ser" evaluativo. La segunda respuesta adopta un punto de vista diferente y argumenta que la persona en cuestión sí tiene una razón para hacer la cosa en cuestión, a saber, la razón de que debería hacerlo.

Recordemos la situación: los principios de *P* difieren de los tuyos. Según sus principios, él no debería hacer *H*. Según tus principios, él debería hacer *H*. Más exactamente, hacer *H* por parte de *P* está en concordancia con tus principios pero no con los suyos. Por lo tanto, tú dices que moralmente *P* debería hacer *H*, tal como el análisis de

Hare te permite hacerlo. La objeción fue que esto implica que *P* tiene una razón para hacer *H* que él no tiene. La réplica que ahora estamos considerando es que cuando dices que *P* debería hacer *H*, tú piensas que tiene una razón para hacer *H*, porque que sea el caso que él debería hacer *H* es para él una razón para hacer *H*, lo advierta o no. Es verdad que *P* no supondrá que tiene una razón para hacer *H*, en especial esta razón. Pero decir que *P* tiene una razón para hacer *H* no equivale a decir que *P* advierta, por sí mismo, que tiene una razón para hacer *H*.

Para que esto resulte más concreto, supongamos que uno de tus principios consiste en no comer carne. Éste es un principio al que tú tienes la intención de adherirte y al que también tienes la intención de que los demás se adhieran. Por lo tanto, según Hare, éste es uno de tus principios morales. Ahora bien, resulta que *P* tiene un principio diferente. *P* está en favor de comer carne en todas las comidas. A partir de tus principios, en conjunción con el hecho de que el filete en el plato de *P* es carne, concluyes que *P* no debería comer el filete. Además, dices que *P* tiene una razón para no comer el filete, a saber, que es carne y que uno no debería comer carne. Tú concuerdas, sin embargo, en que *P* no acepta este principio y en que acepta, por cierto, el principio opuesto. ¿Es esto coherente?

Desde luego, si *P* hubiera aceptado tu principio general, el hecho de que el filete sea carne le brindaría una razón para no comerlo. Pero, dado que no acepta tu principio, ¿cómo puedes suponer que este hecho le da después de todo tal razón? Ciertamente, puedes suponer que esto le da una razón sólo si tú piensas que *P* debería aceptar tu principio antes que el suyo. Por lo tanto, tienes que pensar que *P* tiene una razón para aceptar tu principio. Pues, si *P* tiene una razón para aceptar tu principio, entonces el hecho de que el filete sea carne le daría a *P* una razón para no comerlo, aunque de hecho no acepte tu principio. Pero si *P* no tiene una razón para aceptar tu principio u otro similar, no está claro en qué sentido el hecho de que el filete sea carne le proporcionaría una razón para no comer el filete.

Por lo tanto, cuando dices que *P* no debería comer el filete implicas que *P* tiene una razón para aceptar tus principios o tiene un principio similar al tuyo. Pero Hare no puede siempre suponer que éste sea el caso. Como filosófo poskantiano, Hare concuerda con gran parte de lo que Kant dijo, pero no piensa que existan razones, por encima de la mera consistencia, que favorezcan una elección de principios frente a otras posibles selecciones diferentes. La razón, por sí misma, no impone constricciones tan fuertes a tus principios. Por lo

tanto, como filósofo poskantiano, Hare no puede decir que *P*, que no acepta el principio "No comas carne", necesariamente tiene razones para aceptarlo. Pero si *P* no tiene razones para aceptar ese principio, tal como es posible, dada la posición de Hare no puede decirse que *P* debería aceptar ese principio. Pues sería inconsistente decir que *P* debería aceptar el principio cuando no tiene razones para aceptarlo. Pero entonces no puede decirse que *P* no debería comer el filete, si sus principios no le otorgan ninguna razón para no comerlo y no tiene razones para aceptar principios que difieran de aquellos que acepta efectivamente.

Hare tiene que conceder que es posible cierto caso en el que los principios básicos del hablante estén en conflicto con los del agente. En tal caso, la teoría de Hare permite al hablante decir que el agente debería hacer cierta cosa aunque hacerla no esté en concordancia con los principios del agente. Tal juicio implica que el agente tiene una razón para hacer eso; esto puede ser así en esta situación, sólo si el agente tiene una razón para adoptar un principio similar al principio al que apela el hablante. Pero el sentido de la variante propuesta por Hare de la teoría kantiana es que el agente no puede tener razones para adoptar principios diferentes a menos que haya algún tipo de incoherencia o irracionalidad en sus propios principios. Puesto que no es preciso que haya una incoherencia o irracionalidad tal en los principios del agente, la teoría de Hare puede llevar al hablante a una contradicción. Porque la teoría dice al hablante que puede juzgar a un agente tomando como base los principios del hablante, y que tal juicio implica que el agente tiene razones para aceptar los principios del hablante; aunque el sentido de la teoría de Hare es que el agente puede muy bien no tener razones para aceptar principios que difieren de los principios que efectivamente acepta.

Ahora bien, Hare podría enfrentar con facilidad esta objeción modificando su análisis de la siguiente manera. Podría suponer que "*P* debería hacer *H*" significa algo semejante a "hacer *H* por parte de *P* está en concordancia con los principios morales generales a los que suscribo por la presente y que *P* también acepta o que, al menos, tiene buenas razones para aceptar". Pero esto significaría que no puedes formular un juicio moral, usando el "debería" moral, acerca de alguien que no acepta el principio relevante y que no tiene una razón para aceptarlo. Y esto puede parecer incorrecto también. Pues, tal como señala Kant, parece que podemos formular tales juicios acer-

ca de cualquiera, independientemente de cuáles sean sus principios morales.

Caben entonces dos posibilidades. Una es que la teoría de Hare, aun revisada de la manera sugerida, simplemente no funciona, por lo que nos queda el punto de vista de Sartre según el cual la moralidad no es más que buena fe y consistencia. La otra posibilidad es que hay algo incorrecto con nuestro punto de vista ordinario según el cual podemos realizar juicios con el "debería" moral acerca de cualquiera, sin importar cuáles sean sus principios. Hemos considerado ya en el último capítulo que podríamos inclinarnos a abandonar la idea según la cual nuestros principios morales obligan a todos los seres racionales. Tal como vimos, esta idea es en sí misma un reflejo de ese aspecto de nuestro punto de vista moral ordinario que Kant describió y que los filósofos poskantianos dicen que tenemos que abandonar. No podemos tener moralidad tal como es concebida ordinariamente. Quizá todo lo que podamos tener es el sustituto que nos proporciona Hare, dada la revisión que hemos señalado. En ese caso, la moralidad comienza a adquirir el carácter de principios que son aceptados por un grupo de personas. Nuestros principios morales son obligatorios para quienes los comparten o para aquellos cuyos principios les dan razones para aceptarlos.

NOTA ACERCA DE LECTURAS ADICIONALES

La teoría de R.M. Hare está desarrollada en *The Language of Morals*, Oxford University Press, Oxford, 1952.

Para una versión popular del punto de vista de J.P. Sartre, véase "Existentialism is a Humanism", en Walter Kaufmann, *Existentialism from Dostoevsky to Sartre*, Meridian, Nueva York, 1956. [Existe traducción al castellano: *El existencialismo es un humanismo*, del original en francés.]

VIII

COSTUMBRE Y RELATIVIDAD

1. *El internalismo y el externalismo*

La moralidad da a la gente razones para hacer cosas. Si una teoría ética es adecuada, debe explicar por qué esto es así. En consecuencia, la teoría del observador ideal (por ejemplo) es defectuosa debido a que no dice por qué alguien tiene una razón para hacer lo que moralmente debería hacer. La teoría dice que alguien moralmente debería hacer algo si y sólo si un observador ideal no prejuiciado aprobara que lo hiciera; pero la teoría no dice cómo las reacciones de tal hipotético observador ideal no prejuiciado ofrecen a una persona real típicamente prejuiciada una razón para hacer algo, porque la teoría no dice por qué alguien debería preocuparse por las reacciones de esta persona imaginaria.

La objeción a la teoría del observador ideal es un caso del argumento de la pregunta pendiente discutido en el capítulo 2. La teoría del observador ideal es una forma de naturalismo ético. Toda teoría naturalista iguala la aseveración "*P* debería moralmente hacer *H*" con una aseveración de la forma "Hacer *H* por parte de *P* tendrá las características naturales *C*". Si no nos preocupamos explícitamente por el hecho de que una acción sea *C*, para nosotros será una pregunta pendiente si alguien debería hacer algo que tiene las características *C*, ya que nos preocupamos por lo que la gente moralmente debería hacer.

Es verdad que no siempre estamos motivados a hacer lo que pensamos que es correcto. A veces nos vemos tentados a hacer algo distinto, y a veces simplemente no nos interesa preocuparnos por ello. Pero normalmente nos preocupamos por lo correcto y lo incorrecto y una teoría ética tiene que dar cuenta, de alguna manera, de nues-

tro interés. Las teorías éticas pueden ser catalogadas en términos de las diferentes maneras en que tratan de explicar esto. Algunas dicen que nuestro interés es puramente interno a nuestras creencias morales, mientras que otras dicen que depende, al menos en parte, de las sanciones externas. Por ejemplo, una teoría de la ley divina dice que lo correcto y lo incorrecto derivan de la ley de Dios, que Él impone con sanciones divinas; según este punto de vista, nos preocupamos por lo correcto y lo incorrecto porque nos preocupa si vamos a ir al cielo o al infierno. Por otra parte, el emotivismo trata de explicar nuestro interés por lo correcto y lo incorrecto identificando el pensamiento de que algo es correcto con un tipo de interés y el pensamiento de que algo es incorrecto con el tipo de interés opuesto. El emotivismo considera el interés moral como puramente interno a nuestras creencias morales, mientras que la teoría de la ley divina sostiene que se deriva en parte de las sanciones externas. El emotivismo es un tipo de *internalismo*; la teoría de la ley divina es un tipo de *externalismo*.

La mayor dificultad con el internalismo es que, como lo hemos visto, tiende al subjetivismo, dados los hechos manifiestos de la psicología humana, y los límites manifiestos de la racionalidad. Personas diferentes parecen ser capaces de adoptar diferentes principios morales sin ser irracionales. La consistencia tiene límites, por supuesto; y, al menos según Hare, no cualquier principio que pudieras llegar a adoptar vale como principio moral. Pues, según Hare, no sólo tienes que intentar seguir aquellos principios tú mismo, sino que además tienes que intentar lograr que los otros los sigan. Pero aun en el enfoque de Hare, son posibles diferentes conjuntos de principios morales y si los principios morales de una persona son diferentes de los de otra, no hay ningún sentido en el que pueda decirse que el conjunto de principios morales de una persona es correcto y que el del otro es incorrecto, al margen de su incoherencia o inconsistencia interna, excepto en el sentido de que el hablante puede indicar su concordancia con un conjunto y no con el otro.

Un subjetivismo de este tipo es antiintuitivo porque ordinariamente nos inclinamos a suponer que las intenciones, objetivos, metas, planes y proyectos de una persona son un tipo de cosas y que la moralidad es otra. Ordinariamente, no suponemos que lo correcto y lo incorrecto esté determinado por las decisiones y principios de un individuo particular. Estamos inclinados a pensar que la moralidad tiene una fuente externa, no una fuente interna.

Ahora bien, el externalismo considera esencial para nuestro interés por lo correcto y lo incorrecto, que este interés debería derivarse, al menos parcialmente, de algún tipo de sanciones externas. La teoría de la ley divina dice que estas sanciones incluyen la privación del amor de Dios y el castigo en el infierno. Algunas veces se sugiere incluso que, si alguien no cree en el castigo y la recompensa divinos, no creerá en una ley moral objetiva; y además, aunque pueda creer en una ley subjetiva que él mismo se da, se dice que tal ley subjetiva no puede considerarse una ley moral en ningún sentido ordinario, dado que la persona en cuestión estará en condiciones de pensar que es libre de actuar sin constricciones morales objetivas. "Si Dios ha muerto, todo está permitido." Pero esta línea de pensamiento supone con demasiada rapidez que la única sanción externa posible de la ley moral es la retribución divina. El argumento se concibe mejor como un argumento a favor del externalismo: el internalismo puro hace demasiado subjetiva la moralidad; nuestra concepción ordinaria de la moralidad es una concepción de algo que tiene una naturaleza más objetiva. Normalmente una persona supone que la moralidad trasciende los principios que ella misma ha adoptado; si llegáramos a creer que Hare estaba en lo correcto y que la moralidad no estaba más allá de los principios de una persona particular, tendríamos que concluir que no hay constricciones morales del tipo de aquellas en las que ordinariamente creemos. "Si el externalismo ha muerto, todo está permitido."

2. *La costumbre social y la moralidad*

Supongamos entonces que la moralidad depende, al menos en parte, de las sanciones externas. Si no son divinas, tienen que ser humanas, esto es, tienen que ser sociales. Por lo tanto, la moralidad tiene que depender esencialmente de algún tipo de ley social impuesto por la costumbre y la presión social. Por supuesto, la moralidad no ha de identificarse con la legalidad; no tenemos que decir que la ley moral es lo mismo que la ley que establece el *gobierno*. Por otra parte, no tenemos que suponer que la línea divisoria entre la moralidad y el derecho (civil y criminal) es una línea definida. La cuestión de la relación entre la ley y la moralidad es crucial para toda teoría de la moralidad basada en la costumbre social; pero ésta es una cuestión que podemos dejar a un lado por el momento. Podemos definir vagamente las teorías de la costumbre social como teorías que dicen

que la moralidad deriva de reglas o costumbres que la sociedad impone de cierta manera, una manera que puede ser especificada de modos diferentes en diferentes teorías.

Una teoría exitosa de la costumbre social, en este sentido, tendría que tener dos ventajas obvias *prima facie* sobre el internalismo puro. En primer lugar, debería dar cuenta de nuestra sensación intuitiva de que la moralidad tiene una fuente que es externa a los principios que una persona particular ha adoptado. Lo correcto y lo incorrecto no resulta ser cuestión de elección personal. En segundo lugar, la teoría debería comenzar explicando por qué la moralidad tiene el contenido particular que normalmente suponemos que tiene. Por ejemplo, normalmente suponemos que hay restricciones morales relativas a robar, a mentir, a defraudar y que hay deberes morales relativos a ser honestos, ayudar a los otros y demostrar gratitud cuando se nos ayuda. Es obvio que la sociedad está interesada en tales restricciones y deberes. Las reglas contra el robo, la mentira y el fraude, y en favor de ayudar al prójimo, de ser honesto, etc., están a favor del interés general. Al margen de toda suposición acerca de una moralidad previa, podemos entender por qué podría haber presión social para actuar en concordancia con esas reglas: tal presión serviría a una función social. Si podemos suponer que la moralidad deriva de la imposición social de ciertas costumbres, podemos comenzar a explicar por qué existen los tipos de principios morales que pensamos que existen. Por otra parte, si suponemos que la moralidad es simplemente un reflejo de las decisiones individuales, no podemos ofrecer tal explicación (aunque en ese caso podríamos ofrecer una explicación evolucionista: la tendencia a desarrollar valores sociales promueve la supervivencia de la gente que ha nacido con esa tendencia).

Por lo tanto, en una teoría de la moralidad fundada en la costumbre social, la imposición de la costumbre es básica para la moralidad. La moralidad está constituida por las reglas, sean las que fueren, que la sociedad impone. No hay una moralidad previa. Si la sociedad impusiera diferentes reglas, cambiaría lo que es correcto y lo que es incorrecto.

Pero no tenemos que simplificar excesivamente. La teoría no identifica simplemente que algo sea incorrecto con que exista presión social en contra de ello. Eso sería claramente inadecuado. Decir que la moralidad deriva de reglas que se imponen socialmente no equivale a decir que todo lo consuetudinario sea correcto. La esclavitud, por

ejemplo, puede ser incorrecta, de acuerdo con una teoría de la costumbre social, aunque sea consuetudinaria y no haya presión social contra ella. Pues, aunque la esclavitud misma sea consuetudinaria, puede estar en conflicto con otras costumbres que son impuestas socialmente; este conflicto puede no ser reconocido por los miembros de la sociedad debido a la ignorancia de hecho, la estupidez y el autoengaño. Según la teoría de la costumbre social, si la esclavitud es incorrecta en la sociedad, es incorrecta en virtud de las reglas que están socialmente impuestas con respecto a la gente que no es esclava. Estas reglas pueden efectivamente aplicarse a todos y no sólo a los no esclavos, aunque los miembros de la sociedad no reconozcan esto porque creen falsamente que existe una diferencia relevante importante entre los esclavos y la demás gente y, en consecuencia, piensan en los esclavos como, digamos, bestias de carga antes que como gente.

En el caso de las reglas sociales, como en el caso de las reglas personales, puede haber principios más básicos y principios menos básicos. Dados los principios básicos socialmente impuestos y ciertas creencias fácticas, los miembros de la sociedad impondrán otros principios que ellos derivan de los principios básicos y de las suposiciones fácticas. Si sus suposiciones fácticas son erróneas, sus principios derivados serán incorrectos aun desde su propio punto de vista. Una teoría de la moralidad fundada en la costumbre social puede suponer, en consecuencia, que la moralidad deriva de principios que son impuestos socialmente, sin tener que suponer que lo que se impone socialmente siempre sea correcto. Tal teoría de la moralidad no está comprometida a decir que, si existe la costumbre de la esclavitud, la esclavitud tiene que ser correcta. Sólo está comprometida a decir que, si la esclavitud es incorrecta, es incorrecta con referencia a una u otra costumbre que es impuesta socialmente.

Esto significa que, si no existieran constricciones sociales, la esclavitud no tendría nada de malo, dado que en ese caso no habría nada incorrecto de acuerdo con la teoría de la costumbre social. En un "estado de naturaleza" nada sería correcto o incorrecto, ya que ninguna regla sería impuesta socialmente. Ahora debemos considerar la plausibilidad de esto. Tenemos que imaginar gente que vive aislada, que nunca coopera de manera semipermanente; por lo tanto, no hay nada semejante a una sociedad, ni siquiera en pequeña escala. Entonces tenemos que preguntarnos si pensamos si aún habría corrección e incorrección en tal estado de naturaleza.

Ahora bien, cuando considera esto cuidadosamente, mucha gente piensa que, aun en un estado de naturaleza, no deberías matar a otros ni dañarlos de ninguna manera (excepto en defensa propia, etc.); no deberías robarles; deberías tratar de no engañarlos; deberías tratar de ayudarlos si necesitaran ayuda. Otra gente siente que un estado de naturaleza se asemejaría más a una situación de guerra generalizada de todos contra todos en la cual las constricciones morales no desempeñarían ningún papel efectivo. Si el primer grupo de gente está en lo correcto, ninguna teoría de la moralidad fundada en la costumbre social puede adecuarse a nuestra concepción ordinaria de la moralidad. Pero es difícil trazar una imagen clara de cómo sería un estado de naturaleza; por lo tanto, no es fácil estar seguro de si se piensa que las constricciones morales se aplicarían en un estado así. Además, por cierto, no es *claramente* esencial a nuestra concepción usual de la moralidad que las constricciones morales tengan que aplicarse en un estado de naturaleza. En consecuencia, es útil considerar si hay otras consideraciones relevantes para evaluar las teorías de la moralidad fundadas en la costumbre social.

3. *El relativismo otra vez*

Una dificultad posible de tal teoría es su carácter relativista. Las reglas básicas que son impuestas socialmente en una sociedad pueden diferir de las reglas básicas impuestas en otra. Por lo tanto, de acuerdo con una teoría de la costumbre social, lo que es correcto y lo que es incorrecto no siempre será lo mismo en sociedades diferentes. Sin embargo, nuestra concepción ordinaria de la moralidad no parece permitir este tipo de relativismo.

Al considerar las diferencias en los códigos morales de sociedades diferentes es útil distinguir tres casos distintos: (a) casos acerca de los cuales estamos inclinados a decir: "A la tierra que fueres, haz lo que vieres", (b) casos semejantes al ejemplo de la esclavitud, en los que pensamos que, aun con respecto a su propia moralidad, los miembros de la sociedad no deberían actuar tal como lo hacen corrientemente, y (c) casos que no caen bajo (a) o (b).

En la primera de estas categorías caen las convenciones sociales de buena educación, deferencia y etiqueta. Aunque estas convenciones varían mucho de una sociedad a otra, no nos perturban especialmente. Estas convenciones tienen una función social: ayudar a suavizar las relaciones con los demás, haciendo más predecibles ciertas activida-

des y haciendo más fácil la participación en ellas; hacen posible que uno se relaje en diversas situaciones sociales y reducen la confusión. Pero no importa mucho cuáles sean exactamente estas convenciones, siempre y cuando haya algunas reglas. En consecuencia, estas reglas se asemejan más a las convenciones lingüísticas o a las reglas de tránsito.

En la segunda categoría están las convenciones que tienen que ver con el matrimonio y la familia, así como las que conciernen a lo que se considera perversión sexual, a las actitudes apropiadas ante la homosexualidad o las relaciones prematrimoniales, etc. También están en esta categoría las cuestiones relativas al aborto, el infanticidio y la crueldad hacia los animales. No tendemos a pensar que las convenciones de una sociedad con respecto a estos temas pueden determinar, al menos de una manera simple, lo que es correcto y lo que es incorrecto. Así como la esclavitud es injusta para los esclavos, la institución social del matrimonio y la familia, tal como existe en una sociedad dada, puede ser injusta para las mujeres de esa sociedad. Esto también puede ser verdad en el caso de las costumbres de esa sociedad en cuanto a la perversión, la homosexualidad, el infanticidio, el aborto y la crueldad hacia los animales. En estos casos no pensamos "A la tierra que fueres, haz lo que vieres". Un matrimonio norteamericano que viaje a un país árabe no pensará que el marido tendría que tratar a su esposa según la costumbre local.

A pesar de eso, estos casos pueden clasificarse juntos merced a la manera en que se asemejan al ejemplo de la esclavitud. Puede ser que las reglas socialmente impuestas con respecto a las relaciones entre los hombres sean las reglas con referencia a las cuales es incorrecto el tratamiento que se da a las mujeres en la sociedad; así como las reglas que se imponen con respecto a los que no son esclavos pueden ser las reglas con referencia a las cuales es incorrecta la esclavitud. Pues puede ser que el tratamiento que se da a las mujeres esté basado en suposiciones falsas acerca de las diferencias entre el hombre y la mujer. De manera similar, las restricciones morales relativas al comportamiento homosexual pueden derivarse de creencias falsas acerca de la homosexualidad. También nuestro tratamiento de los animales puede basarse en creencias falsas acerca de las diferencias entre las personas y los animales.

No quiero señalar que no habrá dificultades al usar una teoría de la costumbre social para dar cuenta de nuestros puntos de vista ordinarios con respecto a estos ejemplos. Pero no está claro que tal teoría no pueda

dar cuenta de nuestros puntos de vista. En consecuencia, resulta útil considerar ejemplos de casos que caen bajo la tercera categoría, casos en los cuales las prácticas sociales no parecen basarse en errores acerca de los hechos relevantes.

Considérese una sociedad en la que está permitido el canibalismo. En esa sociedad no hay ninguna regla que se oponga al consumo de carne humana, como sí la hay en nuestra sociedad. Para simplificar el caso, no hay que suponer que crían seres humanos con el propósito de comer su carne, tal como nosotros criamos ciertos animales. Tampoco hay que suponer que cazan gente a la manera en que nuestros ancestros solían cazar animales. Por lo tanto, la gente que se comen es gente que ha muerto natural o accidentalmente, gente que ha sido ejecutada legítimamente, por ser criminal, y gente a la que han matado en guerras. En nuestra sociedad pensamos que es incorrecto comer carne humana aun en estos casos. El problema es si, para quienes pertenecen a esa sociedad caníbal, es incorrecto hacerlo.

Aquí, nuevamente, nuestras intuiciones son poco claras y conflictivas. Por un lado, nos inclinamos a pensar que cuando el canibalismo es una costumbre de una sociedad, los miembros de esa sociedad que comen carne humana no pueden ser censurados por hacerlo. Pero nos sentimos incómodos si decimos que es moralmente incorrecto *de su parte* comer carne humana. Por otro lado, también creemos que su práctica es objetable. Y, ciertamente, si uno visitara esa sociedad sería incorrecto de su parte enrolarse en esa práctica. En consecuencia, definitivamente éste no es un caso de "A la tierra que fueres, haz lo que vieres".

Para tomar un ejemplo diferente, supongamos que hay otra sociedad en la que la gente reconoce la obligación de no dañar a los demás pero no reconoce el deber de ayudar a los demás cuando se necesita ayuda. Cuando alguien sufre un accidente y se enferma, nadie en esta segunda sociedad supone que tiene una razón para buscar un médico y a nadie se le ocurre advertirte si hay una avispa en tu hamburguesa. La cuestión consiste en si, pese a la carencia de una regla establecida socialmente, los miembros de esta sociedad deberían moralmente de ayudar a otro cuando pueden hacerlo con un mínimo o ningún costo de su parte. Y nuevamente éste es un problema difícil de solucionar de manera intuitiva. Nos inclinamos a pensar, por otra parte, que hay algo incorrecto en las reglas de la sociedad. Supongamos que debería haber una regla de ayuda mutua impuesta socialmente. Pero esto no significa que, en ausencia de tal regla, los miembros de la sociedad

tengan, no obstante, el deber moral de ayudar a quien está en la miseria y, dadas las prácticas sociales de la sociedad, no es obvio que ninguno de ellos pueda ser censurado por no ayudar a los otros.

Además, no es obvio que, si visitaras esa sociedad y vieras que alguien necesita ayuda que tú y ningún otro puede proporcionar, tendrías el deber de ayudarlo. Pues, si la situación se invirtiera, ese alguien no supondría que tiene alguna razón para ayudarte y no queda claro que estés obligado a ayudar a alguien que sabes que no te ayudaría si se invirtiera la situación. No habría nada incorrecto en ayudarlo; de hecho, sería bueno de tu parte si lo hicieras. Pero esto no quiere decir que tendrías un deber moral o una obligación de ayudarlo.

Por lo tanto, cuando consideramos los casos más cruciales a favor de una teoría de la moralidad fundada en la costumbre social, encontramos difícil decidir si la teoría concuerda o está en conflicto con nuestras intuiciones. Esto se debe en parte a la vaguedad de tales teorías, tal como las hemos descrito, y en parte a la complejidad de nuestras reacciones. En el próximo capítulo consideraremos una teoría más precisa que supone que la moralidad se funda en una convención tácita entre los miembros de la sociedad, y trataremos de ordenar los diferentes aspectos de nuestras intuiciones acerca de esos casos.

NOTA ACERCA DE LECTURAS ADICIONALES

William Frankena contrasta el internalismo con el externalismo en "Obligation and Motivation in Recent Moral Philosophy", en A.I. Melden, *Essays in Moral Philosophy*, University of Washington Press, Seattle, 1956.

Hay una recopilación útil compilada por John Ladd, *Relativism*, Wadsworth, Belmont, California, 1973.

4. RAZONES Y CONVENCIONES

IX

LA CONVENCIÓN

1. *Las convenciones morales*

Hume dice que algunos de los aspectos de la moralidad, aunque no todos, descansan en la "convención". Existe una convención en el sentido de Hume cuando cada una de varias personas se adhiere a ciertos principios de manera tal que cada uno de las demás también se adherirá a ellos. En mi trato con los demás me adhiero a los principios porque su adhesión a los mismos en su trato hacia mí me beneficia y porque pienso que dejarán de adherirse a ellos en su trato hacia mí a menos de que yo siga adhiriéndome a dichos principios cuando los trato. Por ejemplo, dos campesinos tienen la convención de ayudarse mutuamente a cultivar sus campos. El campesino *A* ayuda al campesino *B* a cultivar su campo de manera que, cuando llegue el momento de cultivar el campo de *A*, el campesino *B* lo ayudará. Cada campesino se beneficia con esta práctica, que depende de la expectativa de ambos de que el otro la continúe.

Hume menciona otras convenciones de este tipo; por ejemplo, aquellas que hacen surgir la institución del dinero. Ciertos trozos de papel pueden ser cambiados por bienes sólo porque a su vez serán aceptados por los demás a cambio de otros bienes. Las convenciones del lenguaje pueden brindarnos otro ejemplo, el cual nos indica que las convenciones pueden ser extremadamente sutiles e incluso imposibles de ser descritas de una manera precisa y explícita por parte de una persona común.

Las convenciones se alcanzan a través de un proceso de regateo implícito y de ajuste mutuo. Dos personas que están remando en un bote ajustarán sus acciones a manera de empujar al mismo tiempo. No importa cuál sea su velocidad en tanto ambos remen a la misma

velocidad. Si uno trata de remar con más rapidez y el otro trata de remar más lentamente, deberá llegarse a algún tipo de arreglo.

Según Hume, entre las convenciones más importantes están las que tienen que ver con la propiedad. Para cada persona es útil que haya un sistema de seguridad respecto de sus posesiones. Este sistema es enteramente convencional, y hasta que se desarrolla no existe una cosa tal como la propiedad. Otra convención importante es la que hace posibles los contratos y las promesas. La convención consiste en que, al usar cierta forma de palabras (u otro signo), una persona se obliga a hacer lo que dice que hará. La obligación de cumplir tus promesas, en consecuencia, deriva de una convención previa, según Hume.

Hume dice que el motivo originario para cumplir las convenciones es "natural" antes que moral, lo cual significa que es un motivo de interés propio. Inicialmente cada persona continúa adhiriéndose a los principios convencionales en su trato con las demás, de manera de que ellas continúen haciéndolo en su trato con ella. Eventualmente se desarrollan hábitos. La acción en concordancia con esos principios se vuelve relativamente automática; sería difícil cambiarla. Las obligaciones basadas en esos principios llegan a parecer naturales y obvias. Según Hume, estas obligaciones "naturales" darán la impresión de que son morales tan pronto como reflexionemos con benevolencia acerca de la utilidad de las convenciones relevantes para la sociedad humana. Ya que, como se recordará, Hume acepta un tipo de teoría del observador ideal. Desde su punto de vista, los juicios morales expresan sentimientos basados en la simpatía.

El propio Hume no piensa que todo lo relativo a la moralidad sea convencional, aunque piensa que gran parte de ella lo es. Sostiene que la simpatía puede llevarnos a aprobar o a desaprobar algunas cosas al margen de nuestras convenciones previas: por ejemplo, aprobaremos la bondad hacia los demás aun en un estado de naturaleza, y, según el punto de vista de Hume, ésta es aprobación moral. No obstante, Hume estaría probablemente de acuerdo en que las *obligaciones* y los *deberes* morales dependen de la convención; de cualquier manera, supondré en lo que sigue que esto forma parte de la teoría de Hume.

Una teoría más extrema que la de Hume tratará como convencional todo aspecto de la moralidad. Por ejemplo, mientras que Hume cree que una simpatía débil hacia los demás está incorporada en la gente, se podría suponer en cambio que la simpatía deriva en sí misma de una convención por la cual la gente concuerda tácitamente respecto

de los demás, al menos en la medida de tratar de sentir simpatía por los demás. Pero no precisamos decidir entre la teoría de Hume y esta versión más extrema.

La teoría de la moralidad que propone Hume, fundada en la convención tácita, es una versión más específica de la teoría de la costumbre social. Tiene varias ventajas. En primer lugar, provee una explicación más específica de la manera en que la moralidad involucra la utilidad social: ciertas reglas se adoptan convencionalmente porque cada persona se beneficia con el hecho de que todos los demás actúen en concordancia con ellas. En consecuencia, esperamos que las reglas sean adoptadas si promueven la utilidad social, en el sentido de que son benéficas para todos.

Tomemos otro ejemplo. Tal como advertimos en el último capítulo, normalmente no suponemos que estés obligado a ayudar a alguien cuando sabes que él no te ayudaría si se invirtiera la situación; sentimos que ayudar a esa persona sería hacer algo que está más allá del llamado del deber, sería un acto de generosidad más que algo que estás obligado a hacer. Sin embargo, esto es justo lo que esperaríamos dada la teoría de Hume. Existen razones de interés propio para que la gente adopte una convención de ayuda *mutua*, pero no existen razones obvias de este tipo para extender esta convención a modo de ayudar a quienes no participan en ella. De tal modo, según la teoría de Hume, no esperamos tener una obligación o un deber de ayudar a quien no nos ayudaría. Por otra parte, la simpatía llevaría a un observador a aprobar que ayudes a esa persona; por lo tanto, dada la teoría de Hume, sería bueno que lo ayudaras aunque no estás obligado a hacerlo.

En el capítulo precedente advertimos nuestra renuencia a censurar a los caníbales por comer carne humana, a pesar de que aborrecemos que lo hagan y que consideramos que sería incorrecto que alguno de nosotros lo hiciera cuando visitara una sociedad que practica el canibalismo. Dada la teoría de Hume, podríamos explicar nuestra propia aversión al consumo de carne humana de la siguiente manera. En nuestra sociedad tenemos una convención tácita según la cual respetaremos a los demás en tanto personas. Según la frase de Kant, queremos "tratar a las personas como fines", como si fuesen sagradas y poseyeran un tipo de dignidad especial. Además, existen distintas formas convencionales a través de las cuales llegamos a expresar nuestro respeto y, en consecuencia, si las personas no son tratadas de acuerdo con esas convenciones, llegamos a ver tal cosa como algo que

degrada la dignidad humana. Por ejemplo, si alguien muere, pensamos que es apropiado hacer un funeral y sepultar el cuerpo, o quizá cremarlo. Dadas nuestras convenciones corrientes, no comeremos el cuerpo. Hacerlo nos parecería un insulto a la memoria de la persona que ha muerto. Ello indicaría una falta de respeto por las personas en tanto personas. Nuestro respeto por la gente y nuestros hábitos convencionales para expresar ese respeto nos conducen automáticamente a rechazar la idea de que podríamos comer carne humana; incluso, hemos llegado a encontrar desagradable la idea misma.

Sin embargo, nuestras reacciones ante los caníbales son complicadas porque son relevantes dos moralidades, la de ellos y la nuestra. Al juzgar la situación podemos apelar simplemente a nuestra propia moralidad: "¡Comer gente es incorrecto!" Pero al juzgar a los caníbales tenemos que tomar en cuenta su moralidad. No podemos simplemente censurarlos por lo que hacen porque su comprensión no es igual a la nuestra. No ven nada incorrecto en el consumo de carne humana y no hay una razón obvia para que lo vean. Esto hace difícil juzgar que es incorrecto *de su parte* comer carne humana. No nos resulta fácil juzgar que los caníbales hacen algo incorrecto. No parece correcto decir que niguno de ellos debería moralmente comer carne humana o que cada uno de ellos tiene el deber u obligación moral de no hacerlo. En el mejor de los casos podríamos decir que no debería ser el caso que comieran carne humana; pero como ya hemos visto, éste no es el mismo tipo de juicio. Desde nuestro punto de vista podemos juzgar sus actos y su situación, incluso su sociedad y moralidad; pero, según parece, no podemos juzgarlos *a ellos*.

Esta explicación se relaciona con un punto mencionado en el capítulo 7. Nos inclinamos a suponer que una persona no debería moralmente haber hecho algo en particular sólo si también podemos suponer que tenía una razón para no hacerlo. No podemos suponer que los caníbales moralmente no deberían comer carne humana a menos que supongamos también que tienen una razón para no comer carne humana. El problema es que estamos suponiendo que no tienen tal razón, porque su moralidad no es igual a la nuestra. Dada esta suposición, podemos formular ciertos juicios morales o evaluativos acerca de los caníbales; por ejemplo, podemos llamarlos "salvajes ignorantes". Pero no podemos decir de ellos, correctamente, que al comer carne humana hacen algo moralmente incorrecto o que moralmente no deberían hacerlo.

2. *Juzgando a los extraños*

Ahora bien, es muy difícil lograr una captación clara de tales ejemplos porque no siempre queda claro cuándo tiene alguien una razón para hacer algo y cuándo no la tiene. Tomemos un tipo de ejemplo muy diferente. Hitler, quien mató a millones de personas, fue un hombre extraordinariamente maligno. En cierto sentido podemos decir que no debería haber matado a esa gente y que lo que hizo fue incorrecto. Pero las siguientes aseveraciones son débiles y en alguna medida extrañas: "Fue incorrecto por parte de Hitler haber ordenado la exterminación de los judíos", "Moralmente Hitler no debería haber ordenado la exterminación de los judíos".

Uno podría suponer que lo que hace que esas advertencias parezcan demasiado débiles es la enormidad del crimen cometido por Hitler contra la humanidad. Mató a tanta gente; ya habría sido incorrecto de su parte haber matado a una sola persona. Decir simplemente que fue incorrecto de su parte haber ordenado el exterminio de los judíos sugiere que *sólo* fue incorrecto; esto es, que es incorrecto sólo de la manera en que es incorrecto el asesinato. Y, dado lo que hizo Hitler, esto es como si uno dijera que, al ordenar la exterminación de los judíos, Hitler hizo una *travesura*.

Sin embargo, la explicación no es completamente satisfactoria. En primer lugar, podemos decir cosas acerca de Hitler que no tengan el mismo tipo de "extrañeza". Aunque sería extraño decir que fue incorrecto por parte de Hitler haber actuado como actuó, no es igualmente extraño decir que lo que Hitler hizo fue incorrecto. De manera similar, no hay nada extraño en la aseveración "Lo que Hitler hizo no debería haber ocurrido nunca". No es extraña en la manera en que es extraño decir "Moralmente Hitler no debería haber ordenado la exterminación de los judíos". Pero si la enormidad de su crimen torna extraña una de esas aseveraciones, ¿por qué no torna extraña la otra?

Otra razón para dudar que la enormidad del crimen, por sí misma, sea la razón de la extrañeza de alguno de esos juicios, es que podemos formular esos mismos juicios acerca de alguien que ha cometido un crimen igualmente atroz, al menos si la atrocidad se mide en términos de la cantidad de personas muertas. Por ejemplo, Stalin fue también un asesino de masas, que ordenó las purgas de la década de 1930 sabiendo que morirían millones de personas. Sin embargo, es posible pensar que Stalin sólo estaba tratando de hacer lo correcto, que odiaba la idea de realizar purgas, que, no obstante, también estaba alarmado

por las consecuencias de no ordenar las purgas porque temía que la revolución estuviera en peligro de ser destruida. Stalin se vio enfrentado a una elección terrible y optó por lo que creyó que era el menor de los dos males. No estoy sugiriendo que ésta sea la verdad acerca de Stalin; probablemente no lo es. Sólo quiero decir que ésta es una visión posible de Stalin. Por supuesto, aun quien adopte una visión tan benévola de Stalin puede suponer que Stalin estaba terriblemente equivocado. Adoptar esta visión de Stalin no significa, ciertamente, condonar las acciones de Stalin. Nunca puede ser correcto ordenar de esa manera la muerte de millones de personas, no importa lo que se espere lograr. En verdad, al adoptar esta visión de Stalin es natural decir que fue incorrecto por parte de Stalin haber ordenado las purgas; Stalin fue moralmente culpable por haberlo hecho. La pregunta interesante es, entonces, por qué no es extraño decir esto acerca de Stalin, de la manera en que es extraño decir lo mismo acerca de Hitler. La diferencia no puede radicar en el gran número de personas muertas, dado que ambos mataron a un gran número de personas. Y, ciertamente, el juicio de que fue incorrecto por parte de Stalin haber ordenado las purgas no es el juicio de que fue una *travesura* de su parte haberlo hecho. Entonces, ¿por qué parece que si dices que fue incorrecto por parte de Hitler haber hecho lo que hizo, estás diciendo algo tan extraño y ridículo como si hubieras dicho que Hitler se portó mal al hacer eso?

Parte de la respuesta tiene que ver con nuestra manera de concebir las actitudes que Hitler y Stalin adoptaron con respecto a sus crímenes, con los principios morales que pensamos que ellos adoptaron, con nuestras opiniones acerca de lo que ellos consideraron que eran razones para actuar. En este aspecto, la actitud de Hitler fue mucho más extrema que la de Stalin. Hitler está más alejado de nosotros de lo que lo está Stalin (o del Stalin que imaginamos en la visión que he esbozado). Hitler está más allá de los límites de una manera en que no lo está Stalin. Hitler no sólo fue inmoral; fue amoral, fue malvado. Stalin fue terrible y también, quizá, malvado; pero no estuvo completamente más allá de los alcances de la moralidad tal como lo he imaginado. Tenemos que pensar que Hitler estaba fuera de los alcances de la moralidad o por lo menos de aquella parte de la moralidad que invocamos al juzgar que fue un hombre malvado.

Al decir que fue incorrecto por parte de Hitler haber ordenado la exterminación de los judíos, estaríamos diciendo que Hitler tuvo una razón (toda razón en el mundo) para no hacer lo que hizo. Pero lo

que es horrible acerca de alguien que hizo lo que él hizo es que pudo no haber tenido tal razón. Si quería exterminar a todo un pueblo, no había razón para que no lo hiciera: esto es lo que resulta tan terrible acerca de él. Por eso suena tan débil decir que fue incorrecto de su parte hacer lo que hizo. Esto sugiere que él tenía una razón para no actuar como lo hizo y sentimos que todo hombre que pudiera haber hecho lo que hizo Hitler tiene que ser el tipo de hombre que no habría tenido una razón para no hacerlo. Tal hombre no está equivocado sino que es malvado.

Por esto suena extraño decir que fue incorrecto por parte de Hitler haber actuado tal como actuó, aunque no es extraño decir que el acto de Hitler fue incorrecto. El juicio de que el acto de Hitler fue incorrecto y el juicio de que nunca debería haber ocurrido no implican que Hitler tuviera una razón para no hacer lo que hizo. El hecho de que sintamos que Hitler no fue el tipo de persona que pudo haber tenido tal razón no socava los juicios acerca de sus *actos* de la manera en que socava ciertos juicios acerca de *él*.

Todo esto es explicable en la teoría de la convención tácita de Hume. Hitler, al igual que los caníbales, está fuera del alcance de nuestra moralidad, aunque en una dirección diferente. Podemos juzgar sus actos haciendo referencia a nuestra moralidad, pero no a Hitler en sí mismo, dado que ello implicaría que fue alguien que reconoció los estándares morales que usamos para juzgarlo. Decir "Fue incorrecto por parte de Hitler" o "Hitler debería moralmente no haberlo hecho" implicaría que Hitler aceptaba las convenciones morales relevantes. Pero sus acciones muestran que no las aceptaba. En consecuencia, estuvo más allá de los límites y fue un enemigo de la humanidad.

Hay otros ejemplos que confirman el mismo punto. Consideremos los juicios que podríamos hacer acerca de marcianos que no sintieran interés por nosotros. Supongamos que estos marcianos no se disuadirían de seguir un curso de acción al reflexionar, simplemente, que ese curso de acción dañaría a algún ser humano. Esos marcianos no considerarían esa reflexión como algún tipo de razón. Para ellos, la reflexión simplemente no les dirá nada contra ese curso de acción. En ese caso, no podemos decir que sería moralmente incorrecto por parte de los marcianos dañarnos.

Esto es estar en desacuerdo con Kant, quien diría que, dado que un marciano es un ser racional, tiene una razón para no dañar a ninguno de nosotros, pues nosotros también somos seres racionales. "El marciano no estaría de acuerdo en que lo dañásemos; entonces, ¿cómo puede

estar de acuerdo en dañarnos?" Kant cree que una reflexión de este tipo puede dar al marciano una razón para no dañarnos. Si Kant estuviera en lo correcto, no habría necesidad de convenciones morales. Podríamos arreglarnos únicamente con la razón práctica pura.

Ahora bien, un defensor de la teoría de la convención tácita de Hume supondrá, plausiblemente, que Kant está equivocado respecto de los poderes de la razón práctica pura. Cuando nos cruzamos con los marcianos por vez primera, es posible que ellos no tengan razones para estar interesados en nosotros y, en ese caso, no hay sobre ellos constricciones morales en su trato con nosotros. Si nos dañan, no es una cuestión de moralidad o inmoralidad, aunque bien podría ser una cuestión de guerra entre planetas. Si resulta que no hay manera de que nosotros los dañemos, de modo que ellos no necesitan preocuparse por nosotros ni siquiera por razones de autointerés, entonces, ellos pueden no desarrollar nunca una moralidad que abarque la nuestra.

Por otra parte, si se desarrolla un conflicto que no responde a su interés ni al nuestro, juntos podemos tratar de arribar a convenciones que reduzcan o eliminen ese tipo de conflicto. Por ejemplo, nosotros y ellos podríamos adoptar la convención de respetarnos mutuamente en tanto seres racionales, la cual involucraría, entre otras cosas, tratar de evitar las acciones que dañarían a otros seres racionales. En ese caso habría una moralidad que abarcaría la suya y la nuestra.

Así es como una moralidad surgiría de un estado de naturaleza, de acuerdo con la teoría de la convención tácita. Antes de que se establecieran las convenciones, no habría cosas tales como lo correcto y lo incorrecto; no tendría sentido juzgar lo que la gente moralmente debería o no debería hacer. Pero una vez que un grupo de personas desarrollara patrones convencionales de acción a fin de evitar conflictos con los demás, sus acciones podrían juzgarse con referencia a aquellas convenciones. Las personas que permanecieran fuera del grupo relevante y estuvieran aún en un estado de naturaleza podrían, no obstante, no ser juzgadas así.

3. *Los aspectos convencionales de la moralidad*

Una razón para pensar que la moralidad ha surgido así, como resultado de una convención, es que ciertos elementos en nuestros puntos de vista morales efectivos parecen reflejar lo que sería el resultado del regateo implícito y de los ajustes mutuos entre personas de diferentes

poderes y recursos. Por ejemplo, considérese un punto al que he aludido varias veces. En nuestra moralidad se piensa que dañar a alguien es peor que no ayudar a alguien. Por eso suponemos que un médico no puede despedazar a un paciente, para salvar a otros cinco pacientes, distribuyendo los órganos del primer paciente de acuerdo con las necesidades. Ahora bien, este principio general acerca de dañar y de no ayudar puede parecer irracional e inmotivado pero adquiere sentido si suponemos que estos puntos de vista morales derivan de una convención tácita que surgió entre gente de diferente riqueza, *status* y poder. Porque, mientras que todos se beneficiarían igualmente con una práctica convencional consistente en tratar de no dañar a los demás, algunas personas se beneficiarían considerablemente más que otras con una convención de ayudar a quienes precisan ayuda. El rico y el poderoso no necesitan ayuda y a menudo están en la mejor posición para ofrecerla; por lo tanto, si se adoptara un fuerte principio de ayuda mutua, ellos ganarían poco y perderían mucho, pues terminarían efectuando la mayor parte de la ayuda y recibirían poco a cambio. Por otro lado, el pobre y el débil podrían negarse a concordar con un principio de no interferencia o no perjuicio, a menos que arribaran también a un acuerdo de ayuda mutua. En consecuencia, podríamos esperar un compromiso, tal como en nuestro ejemplo de los dos remeros que llegan a la velocidad media entre las velocidades a las que cada uno de ellos prefiere remar. En el caso presente, el compromiso esperado involucrará un principio fuerte de no perjuicio y un principio más débil de ayuda mutua, que es lo que ahora tenemos. Si nuestros principios morales no fueran, de esta manera, el resultado del regateo y del ajuste, sería difícil ver por qué supondríamos que hay esta diferencia moral entre dañar y no ayudar; y sería difícil comprender cómo nuestros principios morales podrían ser, de esta manera, el resultado del regateo y del ajuste, a menos que derivaran de algún tipo de convención en el sentido de Hume. Por lo tanto, este aspecto de nuestros puntos de vista morales es una prueba en favor de la teoría de la convención tácita de Hume.

Ahora bien, es importante que la teoría de Hume sea una teoría de la convención *efectiva*. Los deberes y las obligaciones son vistos como si derivaran de convenciones efectivas, no hipotéticas. En consecuencia, la teoría de Hume debe distinguirse de las teorías del acuerdo hipotético que dicen que las reglas morales correctas son aquellas con las que la gente concordar*ía* en ciertas condiciones de igualdad. La explicación que da Hume de la motivación moral requiere su teoría de

la convención efectiva y no funciona con ningún tipo de teoría de la concordancia hipotética. Hume dice que actuamos moralmente, primero, por autointerés y, luego, por el hábito de seguir ciertas reglas convencionales. No podemos explicar de la misma manera por qué alguien estaría motivado a adherirse a principios a los que *tendría que haber* acordado adherirse en una posición de igualdad.

Además, la explicación sugerida de la diferencia moral que advertimos entre dañar y no ayudar, depende de la suposición de que nuestra moralidad se apoya en una convención efectiva entre gente de diferentes poderes y recursos. No es fácil entender cómo podría explicarse este aspecto de nuestros puntos de vista morales suponiendo que las obligaciones dependen de lo que acord*aría*mos en una posición de igualdad. Ya que, en tal posición, parece probable que no estaríamos de acuerdo con nuestros principios morales presentes.

4. *La teoría de la convención tácita y la teoría de Kant*

Finalmente, también debe observarse que la teoría de la convención tácita coincide en importantes aspectos con la teoría de Kant, aunque rechaza una de sus ideas clave. Kant arguyó que tenemos que pensar en los principios de la moralidad como principios que cada uno de nosotros legisla para sí mismo y para los demás, a quienes percibimos como si legislaran también los mismos principios. Es esta segunda parte la que distingue la teoría de Kant de la de Hare. Hare dice que cada persona piensa en los principios de la moralidad como principios que él legisla para sí mismo y para los otros, pero Hare no supone que una persona tiene que pensar que los otros legislan los mismos principios. Y esto plantea una dificultad para Hare (he argüido) cuando consideramos la formulación de juicios acerca de alguien que no comparte los principios relevantes. La teoría de la convención tácita, al igual que la de Kant, ve los principios morales como principios para los cuales la fuente es tanto interna como externa; son legislados por los demás y por uno mismo; y representan los principios de una voluntad general. Kant se equivocó al pensar que esos principios están determinados sólo por la razón y, en consecuencia, se equivocó al suponer que eran universales. Sin embargo, estaba en lo cierto cuando hizo énfasis en su objetividad y su carácter interpersonal. Los principios privados de una persona, que esa persona no toma como compartidos por los demás, no representan un caso normal de princi-

pios morales. En el mejor de los casos, representan un caso límite. La moralidad es esencialmente social.

La teoría de la convención tácita de la moralidad, por lo tanto, no es una versión del externalismo puro. Es una combinación: internalismo más externalismo. De acuerdo con esta teoría, los principios que se te aplican no son simplemente los principios que son aceptados convencionalmente por el grupo circundante, cualesquiera que ellos sean; tú también tienes que aceptar las convenciones. De otro modo, no podrían darte razones para hacer cosas, y el juicio acerca de lo que moralmente deberías hacer o dejar de hacer, o acerca de lo que sería correcto o incorrecto hacer, no podría ser hecho con referencia a aquellas convenciones. En el seno de quienes comparten una moralidad común puede existir una persona amoral. Pero tal persona no puede ser juzgada en términos de los principios de los demás, como no pueden serlo Hitler o los caníbales. Si un marciano que no se preocupa por la vida humana decide vivir en nuestro seno, pero no ve ninguna razón para aceptar nuestras convenciones, no podemos decir correctamente que el marciano está moralmente obligado a no dañarnos (aunque podemos juzgar que sería malo que los marcianos nos dañaran). De manera similar, sería un abuso del lenguaje decir de los criminales profesionales insensibles que es moralmente incorrecto de su parte robar a los demás, o que moralmente no deberían matar a la gente. Dado que no comparten nuestras convenciones, no tienen razones morales para abstenerse de robarnos o de matarnos. (Por otra parte, podemos juzgarlos enemigos de la sociedad y podemos decir que deberían ser capturados y puestos en prisión.)

Las moralidades son sociales. Están definidas por las convenciones de los grupos. Pero tú perteneces a más de un grupo y los diferentes grupos tienen diferentes convenciones. ¿Qué convenciones determinan tus obligaciones morales? Todas lo hacen. Dado que perteneces a un número de grupos diferentes, estás sujeto a un número de moralidades diferentes: la moralidad de tu familia, quizá la de tu escuela, una moralidad profesional (tu "ética comercial"), la moralidad de tu vecindario, las distintas moralidades de los distintos grupos de amigos, la moralidad de tu país y finalmente, quizá, una moralidad limitada que compartes con la mayor parte de la humanidad. Estas moralidades algunas veces estarán en conflicto y darán nacimiento a una situación trágica en la que te verás enfrentado a un conflicto de lealtades. En ese caso, no hay una solución moral clara a tu problema. Tienes que

elegir el grupo que es más importante para ti y actuar según sus convenciones.

Hay un caso límite de la moralidad en el que el "grupo" relevante contiene a una sola persona. En ese caso, una persona estará en condiciones de decir que tiene ciertas obligaciones morales que derivan de sus principios personales y no juzgará que los demás están igualmente obligados (por los principios de dicha persona). Por ejemplo, un pacifista puede pensar que moralmente no debería participar en las guerras, aunque puede no formular el mismo juicio acerca de otras personas. No dirá que es incorrecto de su parte participar en las guerras aunque, ciertamente, pensara que es malo para todos que ellos se involucren en guerras. Y hay muchos *otros* casos en los que una persona impone obligaciones y deberes morales sobre sí misma sin suponer que otras personas están obligadas de manera similar.

Éste constituye un caso límite de la moralidad más que un caso central (aunque puede ser un caso común), porque normalmente concebimos una moralidad como un conjunto de principios que pueden usarse para juzgar a más de una persona y porque, tal como vimos en el capítulo 8, pensamos que la moralidad impone constricciones externas sobre alguien. Sin constricciones objetivas externas no habría una cosa tal como la moralidad, como ordinariamente la entendemos, aun cuando la gente se adhiriera a sus propios principios personales. Si únicamente hubiera moralidades individuales, únicamente conjuntos de principios personales y no convenciones grupales, la moralidad tal como normalmente la concebimos no llegaría a existir.

NOTA ACERCA DE LECTURAS ADICIONALES

La explicación de la convención que ofrece Hume aparece en su *Treatise on Human Nature*, libro 3, parte 2, sección 2.

En *Convention*, de David Lewis (Harvard University Press, Cambridge, Mass., 1969), hay un tratamiento exhaustivo de esta noción y de las nociones relacionadas.

Para otro análisis, véase Gilbert Harman, "Moral Relativism Defended", *Philosophical Review*, vol. 74, 1975.

X

RAZONES

1. *La lógica y las razones*

Tal como hemos visto, muchas teorías éticas consideran que la corrección de un juicio moral que contiene "debería" depende de la manera en que dicho juicio se relaciona con ciertos principios prácticos, sin embargo, las teorías difieren en lo que concierne a los principios que consideran relevantes. Kant, por ejemplo, dice que los principios relevantes son los principios universales de la razón práctica pura. Hare dice que son los principios universalizables que suscribe la persona que formula el juicio que contiene "debería". Sartre dice que son los principios aceptados por la persona acerca de quien se realiza el juicio que contiene "debería". Según la teoría de la convención, los principios relevantes son aceptados convencionalmente por los miembros de algún grupo indicado contextualmente que incluye tanto a la persona a la que se juzga como a la persona que formula el juicio.

Tales teorías también difieren en lo que dicen acerca de la relación requerida, los principios prácticos y el juicio moral. Por ejemplo, Hare sostiene que se trata de una relación lógica que debe explicarse dentro de una lógica de los imperativos. Una idea alternativa y más plausible consiste en que la relación relevante es una relación no lógica más débil del tipo de la que se da entre las razones y los juicios para los que las razones son razones.

Según Hare, los principios prácticos relevantes pueden formularse como imperativos generales; y un juicio que contiene "debería" es correcto en relación con un conjunto de imperativos generales, si el imperativo correspondiente se sigue lógicamente del conjunto de imperativos generales, en conjunción con suposiciones fácticas verda-

deras. Supongamos, por ejemplo, que la persona que está formulando el juicio acepta el principio "¡A quienquiera que sea, devuelve lo que debes si puedes y te piden que lo hagas!" Supongamos, además, que los hechos incluyen (a) que Jones le debe diez dólares a Smith, (b) que Smith pidió que se le devolviera el dinero, y (c) que en este momento Jones puede pagar el dinero. Hare cree que de estas suposiciones podemos derivar el imperativo "¡Jones, paga a Smith los diez dólares que le debes!" Por lo tanto, según el punto de vista de Hare, el juicio de que Jones debería de pagar a Smith los diez dólares que le debe es correcto en relación con ese principio general, dados esos hechos. Esta idea acerca de la relación requerida entre el juicio particular que contiene "debería" y los principios generales, también podría ser aceptada en una teoría de la moralidad kantiana, sartreana, o de la convención. Cualquiera de estas teorías podría decir que un juicio moral que contiene "debería" es correcto si y sólo si es una consecuencia lógica de los principios prácticos relevantes y de los hechos del caso. La diferencia entre las teorías tendrá entonces que ver con la relación entre los principios y los juicios particulares que contienen "debería". Pero la idea de que tiene que darse este tipo de relación lógica entre los principios y los juicios particulares no es muy plausible. Por una parte, implica que los principios relevantes son tan detallados que determinan lógicamente de manera exacta lo que tenemos que hacer en toda circunstancia posible. Hay un aire de irrealidad en la idea de que los principios morales tienen este tipo de precisión.

Además, hay cierta impresión de los juicios que contienen "debería", que esta idea no es capaz de expresar. "*P* debería hacer *H*" no es el único tipo de juicio que queremos formular en relación con los principios prácticos relevantes. También queremos estar en condiciones de formular juicios tales como "Sería bueno que *P* hiciera *H*", "Sería incorrecto que *P* no hiciese *H*", "*P* podría hacer *H*", "*P* tiene que hacer *H*". No todos estos juicios tienen todos el mismo significado. Algunos son apropiados, mientras que otros no lo son. La teoría de Hare parece ser más apropiada para los juicios del tipo "*P* tiene que hacer *H*" que para los del tipo "*P* debería hacer *H*". Porque decir que *P* debería hacer *H* no es necesariamente decir que hacer *H* por parte de *P* es requerido de modo absoluto. Si eso es lo que significa, sería mejor decir que *P* tiene que hacer *H* o que sería incorrecto que *P* no hiciera *H*. Podemos suponer que *P* debería hacer *H* sin suponer que tiene de modo absoluto que hacer *H*, o aun que sería incorrecto de su parte no hacer *H*. Decir que *P* debería hacer *H*, por otra parte,

es más fuerte que decir simplemente que sería bueno por parte de *P* que hiciera *H*.

2. *Un análisis en términos de buenas razones*

Esto sugiere una teoría alternativa de acuerdo con la cual los principios prácticos relevantes no tienen el tipo de precisión legal requerida por la teoría de Hare. Según esta teoría alternativa, los principios relevantes nos comprometen, no tanto a acciones particulares en circunstancias particulares, como a ciertos objetivos y metas. Tenemos que respetar a los demás, tenemos que tratar de no dañarlos; si no nos cuesta mucho, tenemos que tratar de ayudar a quienes necesitan ayuda; etc. También hay principios relativos a los deberes y a las divisiones de las responsabilidades, por ejemplo, dentro de las familias: los padres son los responsables de la educación y del bienestar de los niños, etc. Según esta teoría, los principios morales relevantes no son lo suficientemente precisos como para permitirnos determinar lo que alguien debería hacer, si esto tiene que estar determinado lógicamente de la manera en que lo señala Hare, puesto que no podemos *deducir* lo que debería hacerse, dados los hechos.

Sin embargo, en esta teoría alternativa, como en la de Hare, aceptar los principios como principios prácticos es estar motivado para actuar de ciertas maneras. Es tener ciertas metas y fines además de las metas y los fines que tendrías si no hubieras aceptado aquellos principios. Y así como tus otras metas, fines, deseos y planes pueden darte razones para hacer cosas, también pueden hacerlo las metas y los fines que tienes como resultado de tu aceptación de las convenciones de la sociedad, por ejemplo. Pero decir que esos principios te dan razones para hacer cosas no es decir meramente que esos principios implican lógicamente ciertos imperativos. La relación de las razones es más débil que la relación de implicación lógica *vía* una lógica de los imperativos, o al menos diferente de ella. Es esta relación de razones menos precisa la que, según esta teoría, expresa correctamente la conexión entre los principios prácticos relevantes y los juicios particulares que contienen "debería" y que son correctos, en relación con esos principios, dados los hechos.

En esta teoría, entonces, decir que *P* debería hacer *H* es decir que *P* tiene razones suficientes para hacer *H* que son más fuertes que las que tiene para hacer otra cosa. Si lo que quieres decir es que *P* moralmente debería hacer *H*, quieres decir que *P* tiene razones *morales*

suficientes para hacer *H* que son más fuertes que las que tiene para hacer otra cosa. En otras palabras, dados los principios prácticos relevantes, por ejemplo, los principios que *P* acepta como las convenciones de la sociedad, *P* tiene suficientes razones que se derivan de aquellos principios para hacer *H*, etcétera.

Las razones relevantes pueden tener fuerzas diversas, que dan cuenta de los diversos juicios que podrían formularse: "*P* debería hacer *H*", "*H* es lo mejor que *P* puede hacer", "*P* tiene que hacer *H*", "*P* puede hacer *H*", etc. Por ejemplo, dado que *P* intenta adherirse a ciertos principios y dados los hechos, puede ser que haya buenas razones para que *P* hiciera *H* aunque no fuera para nada erróneo que *P* no hiciera *H*. En ese caso, puede no ser verdad que *P* debería hacer *H* aunque sería bueno que *P* hiciera *H*. Por otra parte, si, dada la intención de *P* de adherirse a esos principios y dados los hechos, sería un error por parte de *P* no hacer *H*, *P* debería hacer *H*. Y si fuera irracional por parte de *P* no hacer *H*, dado que efectivamente intenta adherirse a esos principios, *P* tiene que hacer *H*. En ese caso, no hacer *H* sería incompatible con seguir manteniendo la intención de adherirse a esos principios. Decir, en relación con ciertos principios, que *P* tiene que hacer *H*, es decir que si *P* no hace *H*, eso mostraría que de hecho *P* no intenta adherirse seriamente a esos principios. Finalmente, decir que *P* puede hacer *H* o podría hacer *H*, en relación con ciertos principios prácticos, es decir que la aceptación por parte de *P* de esos principios no vuelve irracional de su parte el hacerlo.

Este tipo de teoría se conoce también como "análisis en términos de buenas razones" de los juicios morales que contienen "debería" y de los juicios morales relacionados. Según esta teoría, tales juicios dicen algo acerca de las razones morales que una persona tiene para hacer algo. Un análisis similar sería inapropiado para otras categorías de juicios morales tales como el juicio de que *P* es moralmente malvado al hacer *H*, dado que tales juicios no dependen, de manera similar, de suposiciones acerca de las razones de *P* para hacer cosas.

Los análisis en términos de buenas razones también pueden ofrecerse para otros sentidos de "debería" y términos relacionados tales como "puede" y "tiene que". Recordemos que la palabra "debería" parece tener como mínimo cuatro significados diferentes. Además del "debería" moral que hemos estado discutiendo está el simple "debería" de racionalidad, como cuando decimos que el asaltante del banco

debería usar la puerta trasera; el "debería" evaluativo, como en "Debería haber más amor en el mundo"; y el "debería" de expectativa, como en "El tren debería estar aquí en tres minutos". Hay diferentes sentidos de "tiene que" y "puede" que corresponden a estos diferentes sentidos de "debería". Hay, por ejemplo, un "tiene que" de expectativa, como cuando decimos "El tren tiene que llegar en este momento" y hay un "puede" o "podría" de expectativa, como cuando decimos "El tren puede llegar en pocos minutos" o "El tren podría haber llegado a mediodía; no estoy seguro". De manera similar, hay un "tiene que", un "puede" y un "podría" de evaluación y también de racionalidad simple. Decimos, por ejemplo, "El huracán no tiene que azotar Miami" o "El asaltante del banco tiene que cortar los cables de la alarma si quiere evitar que lo sorprendan".

Los análisis en términos de buenas razones, cuando son modificados adecuadamente, son apropiados para estos casos diferentes. En el caso del "debería", del "tiene que" y del "puede" de expectativa, lo relevante son las razones de creencia. Decir que el tren debería estar aquí pronto es decir que hay buenas razones para pensar que el tren estará aquí pronto. Decir que el tren tiene que haber llegado al mediodía es decir que las razones para pensar esto son concluyentes. Decir que el tren puede no haber llegado es decir que no es incompatible con las razones que corrientemente tenemos para suponer que el tren no haya llegado. De manera similar, decir que el huracán no tiene que azotar Miami es decir que hay razones abrumadoras para tener la esperanza de que el huracán no azotará Miami. Decir que debería haber más amor en el mundo equivale a decir que hay razones para desear que haya más amor en el mundo. Decir que el asaltante del banco debería usar la puerta trasera es decir que tiene buenas razones para usarla. Decir que tiene que cortar los cables de la alarma es decir que sería irracional de su parte no hacerlo, dados sus fines y dados los hechos.

3. *Los "debería"* prima facie

Una ventaja del análisis en términos de buenas razones es que nos permite dar cuenta de los distintos usos de "debería", "puede" y "tiene que". Otra es que el análisis en términos de buenas razones ayuda a explicar un aspecto de nuestros usos de "debería" en el que hizo énfasis el filósofo inglés W.D. Ross (1877–1970), quien arguyó que había dos "debería" de tipo moral, un "debería" *prima facie* y un "debería" considerado todo lo relevante [*an all-things-considered*

"ought"]. Ross argumentó que si tratamos de enunciar nuestros principios morales usando la palabra "debería", tenemos que usar lo que denomina el "debería" *prima facie*. Por ejemplo, decimos "Deberías cumplir tus promesas", "Deberías decir la verdad", "No deberías dañar a los demás", "Deberías ayudar a los necesitados", etcétera. Pero no suponemos que estos principios sean absolutos. No suponemos que deberías cumplir todas las promesas; concedemos que hay circunstancias en las que puedes romper una promesa —de hecho, en algunas circunstancias deberías romperla. Lo mismo vale para otros principios morales. "Todos los principios morales tienen excepciones."

La reacción de Hare ante esta afirmación es suponer que nuestros principios morales son en realidad mucho más complicados. Para Hare, las excepciones están incorporadas a los principios morales; sin embargo, Ross utilizó un enfoque diferente. Según Ross, un principio como "Deberías cumplir tus promesas" es verdadero tal como está; pero este principio no significa que no haya situaciones en las que puedas romper una promesa. Según Ross, el "debería" utilizado para enunciar este principio es un "debería" *prima facie*. Lo que se quiere decir mediante el principio es que, *prima facie*, deberías cumplir tus promesas. Más precisamente, significa que si has prometido hacer algo, eso te da una razón moral para hacer lo que has prometido hacer. Si no tienes otras razones, entonces tendrías que hacer lo que has prometido hacer. Si tienes otras razones, entonces debes sopesar tus distintas razones a fin de decidir lo que deberías hacer tomando en cuenta todo lo relevante.

Por ejemplo, has prometido asistir a una reunión pero tu tía fallece, por lo que también tienes la obligación de asistir a su funeral. En consecuencia, tienes obligaciones que están en conflicto. Dado que has prometido ir a la reunión, deberías ir (*prima facie*) a la reunión. Dado que tu tía ha muerto y que tus padres te estarán esperando en el funeral, deberías ir (*prima facie*) al funeral. Estos enunciados que contienen "debería" no se contradicen, pues son enunciados de "debería" *prima facie*. Ninguno de los enunciados dice lo que deberías hacer una vez considerado todo lo relevante. Esto sólo puede determinarse sopesando tus razones y decidiendo cuál es más fuerte.

La teoría de Ross de los "debería" *prima facie* tiene mucha relevancia para el punto de vista del análisis en términos de buenas razones. Según este punto de vista, los principios morales indican aquello que haces para lo cual tienes razones morales. Lo que tendrías que hacer, considerado todo lo relevante, no está determinado por ningún

principio moral ulterior, sino que está determinado en cualesquiera de las maneras en que las razones en conflicto son sopesadas, ya sean razones morales o razones de otros tipos.

Dicho esto de otra manera, podríamos tomar la teoría de Ross como una advertencia acerca de la forma lógica de los enunciados que contienen "debería". Hablando estrictamente, podemos decir, un enunciado que contiene "debería" posee la siguiente forma: "En relación con *C*, *P* debería hacer *H*". Por ejemplo: "En relación con el hecho de que has prometido hacer algo, deberías hacerlo". Un juicio que contiene "debería" en el que se considera todo lo relevante tendría entonces la forma siguiente: "Considerado todo lo relevante, *P* debería hacer *H*"; en otras palabras, "En relación con todo lo relevante, *P* debería hacer *H*". Esto equivale a alegar que el "debería" *prima facie* es básico y el "debería" considerado todo lo relevante ha de ser definido en términos del "debería" *prima facie*. Podríamos entonces definir el "debería" básico, *prima facie*, de la siguiente manera: "En relación con *C*, *P* debería hacer *H*" significa lo mismo que "*C* da a *P* una razón para hacer *H*".

Pueden formularse observaciones exactamente similares con respecto al "debería" de expectativa. Hay un "debería" de expectativa *prima facie* y un "debería" considerado todo lo relevante. Decimos: "De acuerdo con el horario, el tren debería de estar aquí en cinco minutos; pero, como el maquinista es nuevo, el tren debería de estar aquí un poco más tarde." Podemos definir la forma básica del "debería" de expectativa de la siguiente manera: "En relación con *C*, debería de ser que *S*" significa lo mismo que "*C* tiene una razón para creer que *S*". De manera análoga, podríamos definir la forma básica del "debería" evaluativo del siguiente modo: "En relación con *C*, debería de ser que *S*" significa lo mismo que "*C* es una razón para desear o esperar que *S*".

Finalmente, consideremos el "debería" de racionalidad simple. Kant sostiene que los requerimientos morales derivan sólo de la razón; según Kant, no actuar moralmente es irracional. Si eso fuera verdad, no habría razón para distinguir el "debería" moral del "debería" de simple racionalidad. Y aun cuando supongamos que Kant está equivocado con respecto a los poderes de la razón, podríamos todavía suponer que el "debería" moral es un uso especial del "debería" de racionalidad: no es que sea irracional no actuar de acuerdo con la ley moral, sino más bien que los juicios morales que contienen "debería" son juicios en los que se usa el "debería" de racionalidad respecto de

los agentes que, se cree, intentan adherirse a los principios morales relevantes.

La forma lógica básica de un enunciado que contiene "debería", cuando usa el "debería" de racionalidad, es "En relación con *C*, *P* debería hacer *H*". Se sugiere que éste es el "debería" moral si las condiciones *C* incluyen la intención de *P* de adherirse a los principios relevantes. En ese caso, al decir "En relación con el hecho de que C^1, *P* moralmente debería hacer *H*" estamos diciendo "En relación con el hecho de que C^1 y que *P* tiene la intención de adherirse a los principios relevantes, *P* debería hacer *H*". De acuerdo con esta sugerencia, la diferencia entre el juicio según el cual el ladrón del banco debería desistir de su plan y el juicio según el cual el ladrón del banco debería usar la puerta trasera consiste en que en el primer juicio, pero no en el segundo, consideramos que las condiciones relevantes *C* incluyen la aceptación de ciertos principios por parte del ladrón. Lo que queremos decir es que, dada su aceptación de esos principios, tiene razones para volver a su casa en vez de entrar en el banco y robar.

Esto es reducir el "debería" moral al "debería" de racionalidad. También se pueden ofrecer reducciones análogas de los otros "debería". El "debería" de expectativa en "En relación con *C*, debería ser que *S*" se transforma en "En relación con *C*, uno debería creer que *S*". El "debería" evaluativo de "En relación con *C*, debería ser que *S*", se transforma en "En relación con *C*, uno debería esperar y desear que *S*".

Aún tiene sentido decir que el término "debería" posee cuatro significados diferentes, dado que las reducciones no son las mismas en uno y otro caso. Una oración de la forma "En relación con *C*, *P* debería hacer *H*" podría significar alguna de cuatro cosas diferentes que pueden expresarse, usando el "debería" de racionalidad, de la siguiente manera: "En relación con *C*, *P* debería hacer *H*", "En relación con *C* y con que *P* acepta las convenciones morales que nosotros aceptamos, *P* debería hacer *H*", "En relación con *C*, uno debería creer que *P* hace *H*", "En relación con *C*, uno debería esperar o desear que *P* hiciera *H*". Éstas no son equivalentes. Tampoco es muy correcto decir que el "debería" moral es apenas un caso especial del "debería" de racionalidad. La diferencia consiste en que, cuando usamos el "debería" de racionalidad para decir que *P* debería hacer *H*, no estamos respaldando, necesariamente, que *P* haga *H*, pero cuando usamos el "debería" moral para decir que *P* debería hacer *H* (normalmente)

estamos respaldando que *P* haga *H*. Cuando digo que el ladrón del banco debería usar la puerta trasera no respaldo que lo haga; no estoy indicando que estoy en favor de que lo haga. Pero cuando digo que el ladrón del banco debería desistir de su plan, respaldo que haga tal cosa; estoy indicando que estoy a favor de que desista de su plan. En consecuencia, el "debería" moral es el "debería" de racionalidad más algo. Cuando uso el "debería" moral, presupongo que el agente y mi audiencia aceptan ciertos principios prácticos que yo también acepto, y formulo mi juicio en relación con esos principios.

Consideremos también las maneras en que reaccionamos al advertir que un agente no tiene las metas que supusimos que tenía. Juzgo que un ladrón de bancos debería usar la puerta trasera porque supongo que su meta es robar el banco y huir sin ser descubierto. Si advierto que no intenta robar el banco, sino que simplemente está efectuando un depósito, entonces cancelo mi juicio según el cual el ladrón debería usar la puerta trasera y digo que yo estaba equivocado. Por otra parte, supongamos que uso el "debería" moral para decir que el ladrón de bancos debería desistir de su plan. Al decir esto, estoy presuponiendo que ciertos principios que el ladrón de bancos acepta le dan razones para dejar de ser un ladrón de bancos. Si advierto que el ladrón de bancos, sin embargo, es totalmente amoral y que, dados sus metas y planes, no tiene ninguna razón para no seguir siendo un ladrón de bancos, entonces, aunque cancelaré mi juicio, no lo haré diciendo que yo estaba *equivocado*. Esta diferencia indica que los dos tipos de juicios no son de la misma clase. El sentido moral de "debería" es distinto del sentido del "debería" de racionalidad, aunque ambos están estrechamente relacionados.

4. *Algunas complicaciones*

Sin duda, hablar del "debería" *moral* es una simplificación excesiva, dado que se usa el mismo sentido de la palabra cuando se dice que alguien tiene razones para hacer algo en relación con las reglas del derecho, las reglas del club, las convenciones de etiqueta, las reglas de un juego, etc., que el hablante considera aceptadas por el agente. Si, por ejemplo, resulta que el agente no acepta esas reglas o convenciones, el hablante no cancela su enunciado originario diciendo que estaba equivocado. El hecho de que "debería" tenga el mismo sentido en todos estos casos es un apoyo adicional para la teoría de la moralidad basada en la convención social. Dada esa teoría, parece apropiado

decir que aquellos que aceptan las reglas del derecho, las reglas del club o las convenciones de etiqueta, las aceptan de la manera en que aceptan las convenciones morales.

Otra complicación es que el "debería" moral puede usarse en relación con una moralidad que el hablante no comparte y, en ese caso, al juzgar que *P* debería hacer *H*, el hablante no respalda necesariamente que *P* haga *H*. Consideremos juicios tales como éste: "En tanto cristiano, deberías ofrecer la otra mejilla; sin embargo, me propongo volverte a golpear." Un espía que ha sido descubierto por un amigo podría decir: "Espero que no me delates, aunque me doy cuenta de que, en tanto ciudadano leal, deberías hacerlo". En tal caso, si resulta que el agente no es cristiano o no es un ciudadano leal, el hablante puede cancelar su juicio originario diciendo que estaba equivocado. La diferencia entre el "debería" moral y el "debería" de racionalidad no surge en ejemplos como éstos; sólo surge en el caso más usual en el que el hablante comparte los principios relevantes y respalda el hacer *H* por parte de *P*.

NOTA ACERCA DE LECTURAS ADICIONALES

La teoría de R.M. Hare se explica en *The Language of Morals*, Oxford University Press, Oxford, 1952; y en *Practical Inferences*, Oxford University Press, Oxford, 1971.

Kurt Baier define un análisis en términos de buenas razones en *The Moral Point of View*, Cornell University Press, Ithaca, N.Y., 1958, cap. 3: "The Best Thing To Do".

W.D. Ross discute la noción de deber *prima facie* en *The Right and the Good*, The Clarendon Press, Oxford, 1930, cap. 2.

Donald Davidson considera la idea de Ross como una intuición acerca de la forma lógica en "How Is Weakness of the Will Possible?" en Joel Feinberg (comp.), *Moral Concepts*, Oxford University Press, Oxford, 1969.

Philippa Foot señala una semejanza entre el "debería" moral y el "debería" de etiqueta, etc., en "Morality as a System of Hypothetical Imperatives", *Philosophical Review*, vol. 81, 1972.

XI

UNA TEORÍA NATURALISTA DE LAS RAZONES

1. *Las razones y el razonamiento*

Si un análisis en términos de buenas razones del tipo que hemos descrito en el capítulo anterior es exitoso, los juicios morales que contienen "debería" pueden explicarse en términos de juicios que usan el "debería" de racionalidad, y estos juicios pueden explicarse a su vez como aseveraciones acerca de razones. Pero esto no produce una reducción naturalista de los juicios morales a juicios acerca de hechos de la naturaleza. Por una parte, hay otros juicios morales además de los juicios que contienen "debería" y los juicios relacionados que utilizan "puede" y "tiene que". Por ejemplo, tenemos el juicio según el cual Hitler fue un hombre malvado, y no hemos dicho nada acerca de la manera en que podría analizarse un juicio como ése (tampoco diremos nada de eso en lo que sigue). Además, aunque un análisis en términos de buenas razones reduce los juicios morales que contienen "debería" a juicios acerca de razones, esos juicios acerca de razones son ellos mismos juicios normativos y, en consecuencia, no son claramente juicios acerca de hechos de la naturaleza. No obstante, si un análisis en términos de buenas razones es exitoso, podemos muy bien sentir que se ha logrado algún progreso.

Las razones probablemente tienen algo que ver con el razonamiento posible de alguien. Podría señalarse que el que alguien tenga (buenas) razones para hacer algo equivale a que hay (buen) razonamiento "a su disposición", el cual lo llevaría a decidir hacerlo. Exploremos esta idea. Entonces, para decir en qué consiste que alguien tenga ciertas razones para hacer algo, tenemos que decir en qué consiste que haya cierto razonamiento "a su disposición". Esto significa que tenemos

que decir algo acerca de qué es el razonamiento y también algo acerca de en qué consiste que un razonamiento esté "disponible" de la manera relevante. Comencemos considerando una explicación inadecuada de estos temas y luego pasemos a una más adecuada.

2. *El razonamiento como prueba o como argumento*

La explicación inadecuada considera que un caso de razonamiento se asemeja a una prueba o a un argumento con premisas, pasos de razonamiento y una conclusión. El razonamiento es más o menos concluyente; por lo tanto, hemos de suponer que hay argumentos de diferentes fuerzas. El argumento más fuerte es un argumento *deductivo*, en el que se garantiza que cada paso es verdadero dada la verdad de las premisas o de los pasos previos de los que deriva mediante un principio de lógica deductiva. Un principio tal dice que *no S* y *S o T* implican lógicamente *T*. En otras palabras, la verdad de *no S* y de *S o T* garantiza la verdad de *T*. Si es verdad que Alberto no viene a la fiesta y también es verdad que Alberto viene o Mabel viene, se garantiza que es verdadero que Mabel viene a la fiesta.

No todo razonamiento es concluyente de este modo; no todo razonamiento es deductivo. Por lo tanto, si un caso de razonamiento se asemeja a una prueba o a un argumento, tiene que haber argumentos no deductivos, o *inductivos*, que tengan pasos que están apoyados pero no garantizados por las premisas o por los pasos previos de los cuales se derivan. Tal paso tendría que ser derivado mediante un principio de *lógica inductiva*. Sin embargo, no hay un acuerdo general acerca de los principios de la lógica inductiva y, por lo tanto, no puedo dar un ejemplo claro de un principio tal. Volveré posteriormente al problema que esto plantea.

Hasta aquí sólo hemos considerado el razonamiento que lleva a aceptar una creencia. También está el razonamiento práctico que lleva a decidir hacer algo. Por lo tanto, también tiene que haber (según este punto de vista) argumentos prácticos de diferentes fuerzas, con conclusiones que sean enunciados de intención o, quizá, imperativos.

En resumen, pues, de acuerdo con esta concepción del razonamiento, vemos que no sólo hay argumentos deductivos sino también argumentos inductivos más débiles, así como argumentos prácticos de diversas fuerzas, con conclusiones que son, digamos, decisiones de hacer algo. La lógica deductiva es, según este punto de vista, la

teoría del razonamiento deductivo. También hay una lógica inductiva, que sirve de teoría del razonamiento inductivo, así como una o más lógicas prácticas que proveen una explicación del razonamiento práctico. En la teoría de Hare, por ejemplo, la "lógica de los imperativos" es una teoría del razonamiento práctico. Así, según este tipo de teoría, el razonamiento está disponible si aceptas las premisas de ese razonamiento, pues en ese caso estás justificado en aceptar también la conclusión del razonamiento, si lo haces razonando a partir de esas premisas a través de los pasos relevantes del razonamiento o la prueba.

Esta explicación del razonamiento es inadecuada porque malinterpreta la relación entre el razonamiento y la lógica. Hay argumentos deductivos y pruebas, pero no hay razonamiento deductivo; no en el sentido de deductivo que maneja el lógico, al menos. Los principios deductivos de la implicación lógica no deben confundirse con los principios del razonamiento o de la inferencia. Hemos discutido el principio lógico según el cual *no S* y *S o T* implican lógicamente *T*. Esto no significa que, si tú crees *no S* y además crees *S o T*, estás autorizado a concluir creyendo *T*. Algunas veces, en cambio, abandonarás una de las premisas de tu argumento. De manera similar, es un principio de la lógica deductiva estándar que premisas inconsistentes implican lógicamente cualquier cosa y todas las cosas; y esto no significa que si descubres que tus creencias son inconsistentes estés autorizado a responder creyendo cualquier cosa y todas las cosas.

El razonamiento no sólo puede conducirte a aceptar nuevas creencias, metas, deseos, planes, etc., sino también a rechazar algunas de tus creencias, metas, planes y deseos previos. En consecuencia, el razonamiento no está bien concebido si se lo toma como un argumento o prueba, con premisas, pasos de razonamiento y una conclusión. Más bien, uno empieza con un conjunto de creencias, planes, metas, deseos e intenciones que modifica como resultado del razonamiento, agregando o sustrayendo de este conjunto como sea más conveniente. Hablar de las razones que uno tiene es hablar del razonamiento que podría hacer y que conduciría a modificar el conjunto total de creencias, planes, metas, deseos e intenciones.

La lógica se ocupa del argumento y de la prueba; no se ocupa de manera directa del razonamiento. Algunas veces la lógica es relevante para el razonamiento porque el argumento y la prueba algunas veces son relevantes. Pero el argumento y la prueba no son lo mismo que el

razonamiento. Un argumento, o una prueba, se asemeja más a una explicación que a una inferencia. El argumento puede ser relevante para el razonamiento porque puede tener una función explicativa y porque a menudo las explicaciones son relevantes para el razonamiento. Al razonar, a menudo tratamos de mejorar la coherencia explicativa de nuestros puntos de vista generales.

Hay argumentos deductivos y pruebas, pero no hay una cosa tal como el razonamiento deductivo. Hay razonamiento inductivo, pero no hay necesidad de ninguna lógica inductiva especial, pues no hay argumentos inductivos de la manera en que hay argumentos deductivos. La lógica relevante al razonamiento inductivo es simplemente la lógica deductiva que es, estrictamente hablando, el único tipo de lógica que existe. La lógica deductiva tiene relevancia para el razonamiento inductivo porque éste es a menudo una inferencia a la mejor explicación y la explicación frecuentemente toma una forma deductiva.

Siguiendo a Aristóteles, podríamos denominar "razonamiento teórico" al razonamiento inductivo. El razonamiento teórico es un razonamiento que se ocupa de las creencias; mientras que lo que Aristóteles denominó "razonamiento práctico" se ocupa del deseo o la intención. Decir que alguien tiene ciertas razones para hacer algo es decir que dispone de un razonamiento práctico que lo llevará a decidir hacerlo. La antigua lógica deductiva es, sin embargo, tan relevante al razonamiento práctico como lo es al razonamiento inductivo o al razonamiento teórico. No hay una lógica práctica especial, no hay una lógica de los imperativos, no hay un silogismo práctico. La lógica es relevante para el razonamiento práctico en la medida en que la coherencia explicativa es relevante. En el razonamiento práctico uno trata de hacer sus planes más coherentes internamente y más coherentes con nuestras metas y deseos. Esta coherencia es como mínimo parcialmente explicativa, pues, entre otras cosas, tus planes especifican la manera en que planeas hacer las cosas, y eso equivale a proponerse explicar la manera en que vas a hacer esas cosas. Tales explicaciones pueden tomar una forma deductiva.

3. *Una mejor explicación del razonamiento*

En consecuencia, la teoría del razonamiento no debe identificarse con la lógica. La teoría del razonamiento, si tal teoría es de algún modo posible, es una materia normativa. La lógica no es en el mismo

sentido una materia normativa aunque, como he señalado, es relevante de cierta manera a la teoría normativa del razonamiento.

Un caso de razonamiento podría identificarse simplemente como una transición que va de un conjunto de creencias, planes, metas, deseos e intenciones antecedentes al conjunto modificado que resulta. En general, la racionalidad no requiere ningún razonamiento particular en un momento dado. Ordinariamente, cualquiera de diversos cambios en tu punto de vista antecedente podría ser aceptable. No estás forzado a razonar sólo de una manera. Podrías razonar de esta, de aquella, o de otra manera.

Esta posibilidad de elección da cuenta del uso práctico de palabras tales como "puede", "debería" y "tiene que". Así, por ejemplo, tienes una razón suficiente para hacer *H* si hay un cambio aceptable en tus puntos de vista de este tipo que incluye llegar a decidir hacer *H* (o seguir intentando hacer *H*). En ese caso, al menos, *puedes* o *podrías* hacer *H*. Algunas veces el razonamiento que involucraría que decidieras hacer *H* (o seguir intentando hacer *H*) es claramente preferible a otro razonamiento a tu disposición, aunque ese otro razonamiento es todavía posible sin que sea efectivamente irracional. En ese caso, *deberías* hacer *H*, aunque no *tienes* que hacer *H*. Por otra parte, si fuera irracional de tu parte no razonar de manera que decidieras hacer *H* (o seguir intentando hacer *H*), entonces *tienes* que hacer *H*; *debes* hacerlo.

¿Qué se requiere para que cierto razonamiento sea un razonamiento aceptable? Para contestar esta pregunta podríamos tener la esperanza de enunciar ciertos principios de razonamiento generales y explícitos, de la misma manera en que estamos en condiciones de enunciar los principios generales explícitos de la lógica deductiva, y así, desarrollar una teoría de las razones rigurosa y normativa. No obstante, es improbable que una teoría semejante sea posible. Existen tantas pruebas de que hay principios de razonamiento generales y explícitos como de que hay principios generales explícitos de apreciación estética.

4. *El razonador ideal*

También podríamos tener la esperanza de realizar una reducción naturalista de los enunciados acerca de razones y del razonamiento, de la manera en que los enunciados acerca de colores se reducen a enunciados acerca de observadores ideales o de la manera en que

los enunciados acerca del ciudadano medio se reducen a enunciados acerca de los ciudadanos en general. Quizá podríamos decir algo como esto: un razonamiento es aceptable si es un razonamiento que podría razonar un razonador ideal; un razonador que funciona idealmente es aquel cuyo razonamiento no está distorsionado o no funciona mal de una u otra manera. Ésta sería una teoría que en cierto sentido estaría a mitad de camino entre la teoría moral del observador ideal y el tipo de teoría que analiza "debería" y "bueno" en términos de papeles, funciones e intereses (que se discutió en el capítulo 2). Sin embargo, hay una objeción a la teoría del observador ideal relativa a los juicios morales que no se aplicaría a esta teoría del razonador ideal; a saber, la objeción de que la teoría del observador ideal no muestra la manera en que la moralidad involucra la apelación a principios. No puede haber una objeción análoga con respecto a la teoría del razonamiento del razonador ideal, porque el razonamiento no apela a principios del razonamiento de la manera en que la moralidad apela a principios morales. (Por supuesto, frecuentemente el razonamiento apela a principios: de lógica, de ciencia, de moralidad, etc. La cuestión es que éstos no son principios del razonamiento.) Por otra parte, para que la teoría del razonador ideal funcione, parecería que también tendríamos que analizar en qué consiste razonar sin distorsión o disfunción, y no está claro cómo podríamos dar ahora un análisis semejante. Por lo tanto, puede parecer poco claro que tal reducción naturalista sea posible.

Pero no necesitamos estar en condiciones de analizar ulteriormente en qué consiste razonar sin distorsión o disfunción. Recordemos que la ética es problemática porque ciertos supuestos hechos morales no parecen desempeñar un papel al explicar por qué alguien realiza las observaciones particulares que realiza. Las observaciones morales se explican adecuadamente, parece, apelando a los principios morales del observador y a su sensibilidad moral, más que apelando a hechos acerca de lo que efectivamente es correcto o incorrecto. Además, alguien que hace algo porque es correcto, hace algo que parece explicarse apelando a la creencia de esa persona de que es correcto, más que apelando a la corrección efectiva de lo que hace. Por otra parte, los juicios acerca de las razones no son problemáticos de la manera en que lo son los juicios morales. Frecuentemente apelamos a supuestos hechos acerca de razones para explicar por qué la gente hace lo que hace o cree lo que cree. Explicamos por qué alguien hace algo o cree algo identificando lo que consideramos son sus razones.

Esas explicaciones en términos de las razones de una persona, presuponen una teoría semejante a la teoría burda de las razones y del razonamiento descrita anteriormente. De acuerdo con esta teoría, hay cierto tipo de proceso psicológico que podría denominarse proceso de razonamiento: llegar a creer algo por ciertas razones o llegar a intentar hacer algo por ciertas razones. La teoría no contiene principios generales explícitos de razonamiento; tampoco parece haber mucha posibilidad de que descubramos de modo inmediato tales principios. Por lo tanto, para ver si ciertas razones o el razonamiento podrían explicar las creencias o las acciones de una persona en particular tenemos que tratar de imaginarnos a nosotros mismos en su lugar, con sus creencias, deseos y principios morales previos, para ver si podemos imaginar a qué tipo de conclusiones podríamos arribar razonando desde esa posición. Es necesario apelar a la imaginación empática porque no podemos apelar a principios explícitos de la teoría del razonamiento para que se nos diga cuáles razonamientos son posibles y cuáles no lo son. No obstante, de esta manera podemos confiar en nuestra imaginación únicamente si podemos suponer que nuestro razonamiento funciona de la misma forma que el razonamiento de otras personas. Tenemos, en consecuencia, que suponer algún tipo de teoría del razonador ideal. Nuestra teoría de las razones y del razonamiento tiene que involucrar la pretensión de que el proceso de razonamiento es universal y es siempre el mismo de persona a persona, excepto cuando es distorsionado o funciona mal de una manera u otra. Ésta es, por supuesto, una hipótesis empírica que podría ser controvertida. Pero es difícil ver cómo podríamos suponer que pudiéramos explicar por qué la gente cree lo que cree o se comporta como lo hace, en términos de las razones que tiene, a menos que podamos suponer que algún tipo de teoría del razonador ideal es correcta. De cualquier manera, la teoría de las razones es una teoría empírica: éste es el punto importante que deberá tenerse en cuenta.

5. *Los hechos morales*

Por cierto que restan algunas cuestiones relativas al *status* científico último de los supuestos hechos acerca de razones, así como cuestiones relativas al *status* último de los hechos psicológicos. Podríamos muy bien preguntarnos, por ejemplo, cómo los hechos alegados acerca de razones se relacionan con los hechos de la física, de la química y de la biología. Pero no hay aquí un problema especial; sólo hay un

problema general relativo a las relaciones entre la psicología y las ciencias físicas.

Por lo tanto, la teoría de las razones y del razonamiento no es particularmente problemática; al menos no es más problemática que cualquier otra parte de la psicología. Y eso significa que la ética puede no ser tampoco problemática, si tomamos en cuenta el relativismo moral. Supongamos que tanto la teoría convencional de la moralidad como un análisis en términos de buenas razones son correctos. Entonces, aunque no hay hechos absolutos de corrección o incorrección, más allá de uno u otro conjunto de convenciones, hay pruebas de que hay hechos relacionados con lo que es correcto o incorrecto, con respecto a uno u otro conjunto de convenciones. Esos hechos serían hechos relativos a si aquellos que aceptan ciertas prácticas convencionales tienen o no, razones para hacer una u otra cosa; y hay pruebas empíricas de que existen tales hechos acerca de razones porque con frecuencia tenemos que apelar a hechos acerca de razones para explicar por qué la gente hace lo que hace y cree lo que cree. Por tanto, hay pruebas empíricas de que hay hechos morales (relacionales).

La idea de que los hechos morales son hechos relacionales acerca de razones es más plausible que el punto de vista kantiano, según el cual los hechos morales son hechos no relacionales acerca de razones, aunque no hemos refutado a Kant. Según el punto de vista de Kant, los hechos morales son hechos acerca de las razones que la gente tiene para hacer cosas; razones que tiene, acepte o no acepte ciertas convenciones. Por supuesto, Kant podría estar de acuerdo en que aceptar una convención afecta las razones que uno tiene; por cierto que también diría que hay razones morales para adherirte a muchas de las convenciones morales de tu comunidad: deberías cooperar con la gente, en igualdad de circunstancias [*everything else being equal*]. Kant incluso podría aceptar la explicación que ofrece la teoría de la convención social en lo que concierne a la diferencia intuitiva entre dañar y no ayudar. Kant podría concordar en que esta diferencia intuitiva surge del hecho de que nuestras convenciones son un compromiso que ha sido elaborado tácitamente entre gente de distintos poderes y recursos; únicamente agregaría que la significación moral efectiva de la diferencia deriva de las razones morales de nuestra sociedad.

Kant también podría aceptar una teoría de razones del razonador ideal; su desacuerdo con la teoría de la convención social es un desacuerdo acerca de los poderes de la razón. Kant sostuvo que todo ser racional tiene razones para abstenerse de dañar a los demás, para ser

honesto y veraz, para ayudar a quienes lo necesitan, etc. De acuerdo con Kant, incluso un marciano o un Hitler o un miembro amoral de la Sociedad Anónima de Asesinatos tiene tales razones (si es un ser racional). Según la teoría de la convención social, por otra parte, esto no es verdadero.

En consecuencia, los dos puntos de vista difieren con respecto a la fuente de las razones morales. Kant ve su fuente en la naturaleza de la razón misma; esto es, en la naturaleza del proceso de razonamiento. En la teoría de la convención social no se considera que el razonamiento en sí mismo posea tal poder: la fuente de las razones morales son los propósitos y metas de la persona involucrada. Pienso que sentimos —al menos yo siento esto— que la teoría de la convención tiene que ser correcta en lo que concierne a los poderes de la razón. Kant pide demasiado a la razón. El peso de la prueba recae sobre Kant; es él quien tiene que mostrar la manera en que la razón podría tener los poderes que le atribuye, y hasta el presente, ni Kant ni ninguna otra persona ha sido capaz de hacerlo. Esto no significa que Kant esté equivocado; aunque, por otra parte, parece poco probable que pueda estar en lo correcto. De cualquier manera, ahora tenemos que intentar ver con más atención las posibles fuentes de las razones morales. Comencemos, entonces, considerando, en el próximo capítulo, el egoísmo.

NOTA ACERCA DE LECTURAS ADICIONALES

Para otro análisis, véase Gilbert Harman, *Thought*, Princeton University Press, Princeton N.J., 1973, especialmente el capítulo 10; y "Practical Reasoning", *Review of Metaphysics*, vol. 29, 1975–1976.

R.M. Hare considera la "lógica de los imperativos" como una teoría de la inferencia en su *Practical Inferences*, Oxford University Press, Oxford, 1971.

5. YO Y LOS OTROS

XII

EL EGOÍSMO

1. *El autointerés*

Hemos visto que ciertos juicios morales son interpretados, de un modo verosímil, como juicios acerca de razones morales. Ahora es tiempo de decir algo acerca de las posibles fuentes de esas razones. Comencemos considerando en detalle una teoría denominada egoísmo, la cual sostiene que las razones prácticas de cualquier tipo, incluyendo las razones morales, derivan en última instancia de consideraciones de autointerés. Según esta teoría, tienes razones para decir la verdad y cumplir tus promesas sólo en la medida en que puedas razonablemente suponer que la honestidad es la mejor política. Tienes que estar en condiciones de suponer que Dios o la sociedad o tu propia conciencia te recompensará por tu honestidad y te castigará por tu deshonestidad. Podrías, por ejemplo, tener alguna razón para pensar que otras personas querrán cooperar contigo de maneras que favorecen tu interés, sólo si las tratas honestamente.

Al principio puede parecer que el egoísmo es una teoría relativamente clara y precisa. Sin embargo, el concepto de autointerés es vago de un modo que introduce un tipo de vaguedad o incluso de ambigüedad en lo que ha de considerarse egoísmo. Por ejemplo, no bastará con decir simplemente que algo favorece tu interés en la medida en que satisface tus deseos, en vez de frustrarlos. El hecho de que desees algo no muestra, por sí mismo, que será de tu interés obtenerlo. Por ignorancia podrías desear algo que no favorece tu interés, porque estás equivocado con respecto al efecto que tendrá. En consecuencia, tenemos que decir, más bien, que algo favorece tu interés en la medida en que satisface, en vez de frustrar, tus deseos *intrínsecos*; y que deseas intrínsecamente algo si lo deseas por sí mismo y no por alguna de

las consecuencias que piensas, posiblemente de manera errónea, que pueda tener.

Pero ni siquiera esto es suficiente para valer como el tipo de egoísmo que quiero discutir. Ciertamente se puede tener un deseo intrínseco de placer; y algunos dirían que sólo el placer puede ser deseado intrínsecamente. La dificultad es que hay gente que supone que también puedes tener deseos intrínsecos por muchas otras cosas. Dirán, por ejemplo, que si estás enamorado de alguien, tienes el deseo intrínseco de que esa persona sea feliz; querrás que esa persona sea feliz al margen de cualquier efecto que su felicidad pueda tener sobre ti. Esto significaría, en términos de la definición de autointerés señalada, que todo lo que promueva la felicidad de otra persona favorece tu "interés" aunque no tenga ningún efecto sobre ti, sea el que fuere; incluso aunque nunca lo sepas. Para mis propósitos, ésta es una interpretación demasiado vaga del autointerés; y el egoísmo definido en términos de ese autointerés constituiría un punto de vista demasiado amplio. Este tipo de "egoísmo" sería compatible con el altruismo más generoso, si tuvieras un deseo intrínseco de hacer feliz a otras personas. Dado ese deseo, tendrías entonces una razón "autointeresada" para promover el bienestar general. No obstante, quiero considerar una idea más restringida del autointerés y una concepción del egoísmo que resultará, consecuentemente, más estrecha. Un egoísta, en este sentido, negará que sean posibles los deseos *intrínsecos* de felicidad de los demás. Un no egoísta, en este sentido, estará de acuerdo en que tus razones para actuar tienen que derivar en última instancia de tus deseos intrínsecos, porque puede creer que puedes tener deseos intrínsecos que no sean autointeresados. Son *tus* deseos, por supuesto, pero eso no los hace necesariamente autointeresados. Hume, por ejemplo, cree que en virtud del poder de la simpatía, algunas veces la gente es capaz de interesarse por los demás de una manera no egoísta y que ese interés puede proporcionarte razones (débiles) para actuar a modo de beneficiar a los otros, al margen de algunas ganancias que esperas obtener en tu favor. En consecuencia, Hume no es un egoísta en el sentido que empleo, aunque él suponga que todas las razones para actuar derivan de tus deseos intrínsecos. Pero no es obvio cómo tiene que definirse la forma más restringida de egoísmo que quiero considerar.

Podríamos tratar de explicar el autointerés en términos de la felicidad, diciendo que algo favorece tu interés en la medida en que promueve tu felicidad y disminuye tu infelicidad. Pero, ¿qué es la felicidad? Si la felicidad consiste en obtener lo que quieres, esta ex-

plicación del autointerés se reduce a nuestra explicación anterior. Si es posible tener un interés intrínseco por otras personas, el bienestar general podría constituir una parte de tu felicidad, al margen de cualquier efecto que tuviera sobre ti, dada una concepción de la felicidad artificialmente amplia. Necesitamos una concepción más estrecha.

2. *El hedonismo egoísta*

¿Cómo tiene que definirse, entonces, la felicidad? Se dice a menudo que la felicidad consiste en el placer y en la ausencia de dolor. Por lo tanto, la versión más popular del egoísmo (y la única versión que consideraremos) es el hedonismo egoísta, la teoría según la cual todo deseo es, en última instancia, el deseo de tu propio placer. Según este punto de vista, el placer —tu propio placer— es lo único que deseas intrínsecamente. No puedes desear instrínsecamente la felicidad de otro; en el mejor de los casos, puedes desear intrínsecamente el placer que obtendrías si ese otro fuese feliz, ya sea porque será bueno contigo o porque obtendrás placer pensando que es feliz.

Tal teoría gana cierto monto de plausibilidad a partir de las consideraciones de la teoría de la motivación y del aprendizaje. Un recién nacido es un hedonista total; sólo busca obtener placer y evitar el dolor. Cuando crece continúa buscando el placer pero aprende a diferir la gratificación. En vez de gozar ahora de un poco de placer, el niño aprende a aspirar a mayor placer futuro. Los padres educan y socializan al niño recompensando algunas de sus acciones y castigando otras. Esto funciona porque el niño desea las recompensas placenteras y quiere evitar los castigos dolorosos. Como resultado de este entrenamiento paterno, el niño llega a desear cosas nuevas, no intrínsecamente, por supuesto, sino debido a que ha aprendido que están conectadas con el placer y la ausencia de dolor. El niño llega a querer comportarse de ciertas maneras, no porque llegue a tener deseos intrínsecos de comportarse así, sino porque ha sido recompensado con experiencias placenteras por tal comportamiento y ha sido castigado con experiencias dolorosas por otro comportamiento. Como un dividendo adicional está, además, el placer de la anticipación: pensar en ciertas recompensas es en sí mismo placentero; así, los pensamientos placenteros se asocian directamente a las acciones por las que el niño ha sido recompensado. El niño puede incluso repetir esas acciones sólo por ese placer, mucho tiempo después de que sus padres hayan dejado de recompensarlo.

Hay, es cierto, ejemplos que aparentemente están en conflicto con el hedonismo egoísta en tanto teoría de la motivación de los adultos. Pero esos ejemplos no son concluyentes. Un caso extremo es el de alguien que se suicida a fin de "devolver el golpe" a otros haciéndolos sentirse mal. Esto presenta un problema, pues es difícil entender cómo una persona que se suicida podría esperar obtener algún placer del acto en sí mismo. Los psicólogos explican la dificultad suponiendo que quien se suicida por este tipo de razón tiene que suponer inconscientemente que de alguna manera va a andar merodeando tiempo después para gozar con el sufrimiento de la otra persona e invocan suposiciones inconscientes de este tipo para explicar otros ejemplos aparentemente contradictorios. Dicen, por ejemplo, que la gente hace testamentos, en parte porque obtiene cierto placer, antes del evento, al pensar en la manera en que ha arreglado las cosas, en parte por el placer que puede obtener mostrando el testamento a aquellos que están mencionados (o no mencionados) en él, y en parte porque imagina, irracionalmente, que de alguna manera se percatará de lo que ocurra después de muerta y, después de todo, se verá influida por lo que suceda entonces.

Según el hedonismo egoísta, uno espera ser recordado luego de su muerte, porque piensa en la muerte como en un tipo de vida. En primer lugar, te gusta que los demás piensen en ti mientras vives. Te da placer saber que los demás están pensando en ti, y la fama tiene beneficios indirectos: la gente gusta de hablar contigo, te invita a sus fiestas. Irracionalmente, supones de manera inconsciente que el recuerdo póstumo y la fama serán así. Imaginas, irracionalmente, que aún estarás allí para beneficiarte del recuerdo y la fama póstumos.

Tal como se señaló en el capítulo 5, podemos invocar de manera similar la teoría freudiana del superyó. Cuando eras pequeño, tus padres te recompensaban si actuabas de ciertas maneras y te castigaban si actuabas de otras. Después de algún tiempo aprendiste a imaginar lo que dirían tus padres acerca de los diversos cursos de acción. Los imaginas desaprobando algunos cursos de acción y aprobando otros. Luego haces lo que piensas que ellos aprobarían y finges que te están recompensando. Esta simulación te da placer, aun cuando sólo sea una simulación. Finalmente, actúas de manera rutinaria a modo de recibir el premio imaginario y las recompensas de un padre ficticio, tu superyó, aunque no te percates conscientemente de que ésta es tu razón.

3. *Una objeción estándar*

Sin embargo, hay una objeción estándar a toda forma de hedonismo egoísta que sea defendido mediante la apelación a creencias y motivos inconscientes. La objeción consiste en que, de defenderse así, el egoísmo se torna incontrastable y no científico; si el hedonismo egoísta es defendido de tal manera, no podemos siquiera imaginar un caso de comportamiento humano que no pudiera explicar, lo que significa que no puede hacer ninguna aseveración empírica con respecto al comportamiento humano.

Consideremos, a modo de ejemplo, un caso de heroísmo en tiempo de guerra.

> Un grupo de soldados está atrapado detrás de las líneas enemigas. Uno de ellos se ofrece como voluntario para correr a descubierto y atraer el fuego enemigo, de modo que los demás soldados puedan escapar. El voluntario cree que, casi con seguridad, lo matarán. Además, es un ateo que no cree en la vida eterna. También es un realista extremo; no supone, por ejemplo, que luego de que lo maten su alma flotará por el campo de batalla a modo de presenciar la huida de sus amigos y oírlos alabar su noble acción.

Si este caso acaeciera, ¿refutaría el hedonismo egoísta? ¡Obviamente no! En primer lugar, el soldado puede pensar que luego no podría vivir en paz consigo mismo si no hiciera lo que prometió hacer. En consecuencia, puede sacrificarse pensando que de esta manera evitará el dolor futuro. Pero, entonces, incorporemos a nuestra descripción del caso que esto no sea verdad.

> El soldado advierte perfectamente bien que no sufriría mucho si en lugar de correr a descubierto a fin de distraer el fuego enemigo, se detuviera y se escondiera detrás de los árboles próximos, de modo que el enemigo localizara a sus amigos y él pudiera escapar durante la batalla, aunque sus amigos resultaran muertos. Posteriormente sentiría arrepentimiento por esto, pero pese a ello estará en condiciones de gozar de la vida. La memoria es misericordiosa. Él sabe que no es el tipo de persona capaz de refugiarse en tales cosas. Sabe que no sufrirá mucho si los engaña y deja que los otros soldados, y no él, sean muertos. No obstante, sabiendo esto, corre a campo traviesa, atrae el fuego enemigo, y muere mientras sus amigos escapan.

¿Refutaría esto el hedonismo egoísta?

No. Quien aboga por esa teoría siempre puede postular, como último recurso, que a pesar del profesado ateísmo del soldado y su creencia consciente de que no hay una vida ulterior, está suponiendo *inconscientemente* que luego estará merodeando por allí para ver y oír la huida de sus amigos y oírlos alabarlo. Inconscientemente, el soldado piensa que obtendrá mucho placer de esta acción, pues piensa inconscientemente que estará merodeando luego del evento. También está el placer que irracionalmente imagina que obtendrá de su superyó como recompensa: aquí está involucrada una doble ilusión. Su superyó es un padre imaginario y, de cualquier modo, ni él ni su superyó estarán allí luego del evento. Además, puede exagerar inconscientemente lo mal que se sentiría si no hiciera lo que prometió hacer.

El hecho de que el hedonismo pueda absorber tales casos no es obviamente un punto a su favor. Por el contrario, parece mostrar que la teoría no puede hacer ninguna aseveración empírica respecto del comportamiento humano. Las cosas serían diferentes si la teoría predijera, por ejemplo, que sólo quien no fuera ateo actuaría de la manera descrita. Entonces, la teoría predeciría que, en la situación descrita, un ateo no correría efectivamente al descubierto para atraer el fuego enemigo, sino que se escondería detrás de los árboles y dejaría que los enemigos mataran a sus amigos. Una teoría que hiciese tal predicción sería una teoría empírica interesante que podría ser contrastada mediante el examen de casos de heroísmo. (La teoría probablemente sería refutada.) Pero un hedonismo egoísta que se defiende apelando a las creencias irracionales inconscientes no puede contrastarse de esta manera. Por lo tanto, parece que no hace aseveraciones empíricas interesantes acerca del comportamiento humano, pues parece compatible con cualquier comportamiento humano imaginable, sea el que fuere. Esto, de cualquier manera, es una objeción estándar contra el hedonismo egoísta, cuando se lo defiende de la manera que he señalado.

4. *Réplica a la objeción*

Sin embargo, la objeción no está bien elaborada. Está basada en la suposición de que el contenido empírico de una pretensión descansa por completo en sus consecuencias contrastables. El problema con esa suposición es que considerará que casi ningún enunciado científico es empíricamente significativo. Muy pocos de esos enunciados son contrastables independientemente de otras suposiciones. Un enuncia-

do singular acerca de la masa de los electrones, por ejemplo, no tiene consecuencias verificables, considerado en sí mismo. El enunciado no hace aseveraciones acerca del comportamiento observable de los macroobjetos. Estamos en condiciones de obtener pruebas a favor o en contra de las hipótesis que se refieren a las partículas y acerca de los instrumentos de medición. Si todas estas suposiciones, tomadas en conjunto, nos permiten realizar una predicción empírica acerca de lo que vamos a observar, y esa predicción no se cumple, podemos decidir que nuestra suposición acerca de la masa del electrón es incorrecta, o podemos en cambio abandonar una u otra de nuestras suposiciones. Si así lo quisiéramos, podríamos continuar aceptando consistentemente cualquier hipótesis acerca de la masa del electrón, en la medida en que deseáramos hacer otras suposiciones para dar cuenta de las observaciones, aunque hacerlo no siempre resultaría lo más racional.

La cuestión es que no evaluamos las hipótesis individuales a la luz de las pruebas empíricas; evaluamos teorías completas. Aunque el hedonismo egoísta tomado en sí mismo no tiene consecuencias contrastables, puede ser una parte útil de una teoría mayor, que sea en sí misma fructífera. Y cuando el hedonismo egoísta es tratado así, como una hipótesis científica —en otras palabras, cuando es considerado una suposición en la teoría—, entonces, para decidir si es verdadero o no (es decir, si queremos aceptarlo o no) no debemos limitarnos a preguntar si hay alguna manera directa de contrastar esa hipótesis particular; en lugar de ello, tenemos que comparar la teoría de la que el hedonismo egoísta se considera parte, con otras que no lo involucran.

Al concebir el hedonismo egoísta como una parte de una teoría más extensa —tal como la teoría freudiana— podemos comprender sus méritos. Por una parte, tal como hemos visto, ofrece una explicación simple de la manera en que se desarrolla la motivación. La apelación a las creencias y deseos inconscientes que tiene que hacer esa teoría es, además, realmente un punto a favor más que en contra. La gente *está* motivada de maneras inconscientes que no comprende completamente y que puede no reconocer. A menudo está atrapada en patrones de comportamiento irracional que se repiten constantemente una y otra vez, patrones que a veces tienen sentido sólo si se supone que el agente está actuando respecto de otras personas como si fueran sus padres o hermanos. Pero eso equivale a decir que podemos dar sentido a gran parte del comportamiento que de otra manera parecería irracional, sólo si atribuimos ciertas creencias y deseos inconscientes a la gente.

Nuevamente, al considerar la manera en que el hedonismo egoísta explicaría los suicidios y los actos de heroísmo, vimos el modo en que podría apelar a la creencia inconsciente de que el agente iba a estar merodeando luego de su muerte. Hay pruebas de que esto es correcto. Por ejemplo, cuando el agente decide hacer algo para herir a otro, el agente piensa a menudo, con muchos detalles imaginarios, la manera en que se va a sentir la otra persona, y al imaginar tal cosa, el agente se imagina a sí mismo como si lo supiera, incluso en los casos en los que se percata enteramente de que nunca lo sabrá.

La teoría del superyó también tiene mucho a su favor. La gente tiende a hablar consigo misma —se censura, se alienta, se alaba, etc.— de maneras que podrían ser apropiadas para los padres o para quienes los sustituyan. Tal como señalamos en el capítulo 5, es un hecho conocido que los atletas tienden a internalizar a sus entrenadores de manera que se dan órdenes a sí mismos, a veces incluso en voz alta, como haría un entrenador severo. Además, la autoridad del entrenador suele explicarse mejor si se supone que el entrenador ha reemplazado de alguna manera al padre. Se actúa hacia el entrenador de la manera en que se actuaría hacia un padre idealizado.

5. *Lo que la gente quiere*

Sin embargo, no estamos interesados en saber si la teoría freudiana, o alguna otra teoría similar, es aproximadamente correcta, sino que estamos interesados, más específicamente, en el hedonismo egoísta. He descrito como freudiana la teoría que considera que la fuente básica de motivación es el deseo del placer y de la ausencia de dolor. De hecho, esto es verdadero sólo para la primera teoría freudiana, basada en el principio del placer. Más tarde, Freud llegó a pensar que el hedonismo egoísta no ofrecía una explicación plausible de ciertos tipos de comportamiento destructivo y, entonces, postuló la existencia de una segunda tendencia, *Thanatos*, una especie de deseo de muerte. La relevancia de esto es que podemos conservar gran parte de la teoría freudiana, o de alguna teoría relacionada, al mismo tiempo que variamos las suposiciones acerca de las tendencias básicas o de las fuentes de motivación. Lo que queremos comparar (entre otras cosas) son teorías que se asemejan básicamente a la teoría de Freud en cuanto a que apelan a la motivación inconsciente, a figuras paternas, a superyós, etc., pero que realizan diferentes suposiciones acerca de las tendencias básicas y de los deseos intrínsecos. En consecuencia,

Freud mismo planteó dos posibilidades, el hedonismo egoísta y su teoría dualista posterior.

Otros psicólogos han señalado otras tendencias básicas posibles, tales como la curiosidad. Así, es posible que un bebé nazca con el deseo innato de saber cosas, aparte del placer que puede recibir de ese conocimiento. Un freudiano podría decir, por otra parte, que la curiosidad es aprendida: el niño es recompensado por aprender cosas y generaliza esto, quizá añadiendo recompensas imaginarias. No es obvio cómo podríamos decidir entre estas teorías. Una rata que nunca ha sido recompensada por correr a través de un laberinto, explorará sin embargo cualquier laberinto en que se la coloque. ¿Muestra esto que la rata tiene una tendencia innata a conocer cosas? ¿Es innata su curiosidad? El freudiano —o el skinneriano— podría decir que lo que es innato no es la curiosidad, sino una conexión entre la obtención de ciertos tipos de conocimiento y la obtención de placer; la rata y el bebé son curiosos porque obtienen placer de su aprendizaje de cosas nuevas. ¿Cómo podemos decidir entre estas teorías? ¿Cómo está hecha la gente? La evolución ha trabajado a manera de que la gente se adapte al mundo que la circunda. Para que esto se produzca, la evolución ha asegurado que seamos curiosos. Pero ¿ha hecho tal cosa incorporando una tendencia especial a la curiosidad o la ha hecho tornando placentera la adquisición de conocimiento? y ¿hay una diferencia efectiva entre estas aparentes alternativas? Éstas son preguntas difíciles y no está claro que la filosofía pueda responderlas.

El bebé recién nacido tiene deseo de comida en ciertos momentos y deseos de beber. La evolución ha diseñado al bebé de manera que cuando su cuerpo esté en cierto estado estará hambriento y cuando su cuerpo esté en un estado diferente estará sediento. ¿Cómo funciona esto? ¿Hemos de suponer que el hambre y la sed son tendencias básicas puestas en acción, en el bebé, por los estados corporales relevantes? ¿O hemos de suponer que el bebé está motivado a sentir placer al llenar su estómago de cierta manera y que está motivado a dar término a la disconformidad que siente cuando tiene hambre y sed? Cualquiera de estos mecanismos parece posible; no es obvia la manera en que podríamos descubrir cuál es el mecanismo real.

No es fácil decir si el hambre es un deseo de comer o un deseo del placer de comer. Pero hay deseos que no parecen de ninguna manera obvia deseos del placer que acompaña a su satisfacción. De hecho, esos placeres parecen depender de un deseo previo. Un hombre obtiene placer, por ejemplo, de la felicidad de sus hijos, porque quiere

que ellos sean felices. Normalmente, no es verdad que quiera que sean felices porque eso lo hace feliz. Aunque eso lo hace feliz, su felicidad depende de su deseo previo de la felicidad de sus hijos. El objeto de ese deseo es la felicidad de sus hijos, no la suya propia. Su propia felicidad no es aquello de lo que su deseo es deseo. O así podría parecer. El hedonismo egoísta tendrá que considerar esta apariencia como ilusoria.

El amor y la amistad también parecen ser cosas que la gente desea en sí mismas, más allá del placer involucrado. Consideremos a alguien que imagina que tiene varios amigos íntimos que lo quieren mucho y una espléndida esposa e hijos que lo aman aunque en realidad los hijos no sean suyos, su esposa lo desprecie y a los que denomina amigos lo toleran únicamente porque se les paga para que lo hagan. Él nunca descubre esto. Vive hasta los 75 años y muere sin tener la menor sospecha. Muere feliz. ¿O no? Su vida, definitivamente, parece haber carecido de algo muy importante, algo que él quería mucho y que pensó que tenía. Esto parecería indicar que lo que una persona quiere puede trascender sus placeres y sus experiencias subjetivas. La mayoría de las personas preferirían una vida en la que sus amigos realmente las admiraran y en la que sus esposas o esposos realmente las quisieran. Preferirían una vida así a una vida que fuera subjetivamente la misma y en la cual, no obstante, no hubieran sido realmente amadas y admiradas, sino despreciadas secretamente. Esto sugiere que los verdaderos intereses de una persona no deberían identificarse con la búsqueda del placer, porque la gente se interesa en más que eso. Al parecer, la gente desea algo más *para ella misma* que el placer; a menudo el placer que tiene es resultado de su creencia de que se han satisfecho sus deseos de otras cosas. Según el hedonismo egoísta, estos deseos son irracionales; pero, ¿por qué tenemos que estar de acuerdo con esto?

Genuinamente, la gente no parece ser capaz de tener una motivación que no sea autointeresada. El hedonismo psicológico —en realidad, toda forma de egoísmo— tiene que tratar todos esos deseos como irracionales y necios; como los de un avaro que ha olvidado el objeto de tener dinero y que ha llegado a desear el dinero por sí mismo; rico pero miserable calcula su riqueza, preocupándose constantemente de que se la roben. Pero el hecho de que a veces sea irracional actuar siguiendo motivos que no sean autointeresados no significa que siempre lo sea. Hay otros casos en los que consideramos elevadamente a una persona que está motivada por lo que podrían parecer consi-

deraciones distintas de las del autointerés. Un bombero salva a una familia que está en un edificio en llamas, pero muere entre ellas. Un prisionero se niega a hablar al ser torturado, aunque resulta claro que favorecería su interés hacerlo. Un soldado se ofrece como voluntario para una "misión suicida" con el fin de salvar a sus amigos. Admiramos a tales personas e incluso esperamos ser capaces de actuar de esa manera si alguna vez las circustancias lo demandaran. De acuerdo con el egoísmo, cuando pensamos así, nos engañamos.

Hay muchos ejemplos de este tipo que no involucran ningún tipo de heroísmo. Un extranjero te pregunta cómo llegar al Nassau Hall. Se lo dices, aunque hacerlo no te interesa particularmente. Parece correcto decírselo e incorrecto no hacerlo, o decirle algo falso. Normalmente supondríamos que tienes alguna razón para decirle dónde está el Nassau Hall; no parecería que esta razón derive de tu autointerés. El egoísmo tiene que negar que haya una razón tal.

Si el hedonismo egoísta es verdadero, sólo tienes una razón para hacer lo que satisfaga mejor tus deseos de placer y de ausencia de dolor. Pero si el hedonismo egoísta no es verdadero y hay otras cosas que deseas por sí mismas, entonces parecerá que tienes una razón para hacer lo que satisfaga esos otros deseos. Esos otros deseos podrían muy bien involucrar la felicidad de otra gente. Alguien te pregunta cómo llegar al Nassau Hall: tienes un deseo intrínseco de ayudarlo; por lo tanto, tienes una razón para decirle cómo llegar al Nassau Hall. El bombero advierte que hay niños atrapados en el edificio en llamas: experimenta inmediatamente un deseo intrínseco de que sean salvados; por lo tanto, tiene una razón para ingresar al edificio y bajarlos desde la ventana del segundo piso aunque esto involucra poner en riesgo su propia vida.

El egoísta quiere saber cómo podemos tener una razón para actuar en el interés de otro si esta razón no deriva del autointerés. La respuesta es que puedes tener tal razón si te preocupas por esa persona en sí misma: puedes tenerla, si tienes un deseo intrínseco de que esa persona sea feliz.

El egoísta considera que es obvio que hay razones para hacer lo que favorece tu autointerés y se pregunta cómo puede haber razones de algún otro tipo, razones altruistas, por ejemplo. Pero, tal como lo han observado los filósofos, si es enigmático que pueda haber altruismo, es igualmente enigmático el autointerés. ¿Por qué tendríamos que preocuparnos por los demás? Bien, ¿por qué tendrías que preocuparte por lo que te pasa? Una persona puede no preocuparse, por una u

otra razón, ni por lo que le pasa a otros ni por lo que le pasa a ella misma. En ese caso, no se puede apelar a las razones de altruismo o de autointerés. Esto suele ser verdad con respecto a las razones autointeresadas a largo plazo de la mayoría de la gente. Tienes una razón derivada de tu autointerés a largo plazo de no fumar; sin embargo fumas, no te preocupas por tu estado dentro de 50 o de 20 años. Las razones derivan de tus deseos, de aquello por lo que te preocupas. Si te preocupas por tu propio futuro, tienes razones autointeresadas para actuar de maneras diversas. Si no te preocupas por tu propio futuro, puedes no tener ninguna de esas razones autointeresadas. Si te preocupas por los demás, tienes otras razones que consideran a los demás, para actuar de maneras diversas; si no te preocupas por la gente, puedes no tener tales razones para actuar de esas maneras.

¿Cómo puedes tener razones que no deriven del autointerés? Tal como Hume creyó, puedes tener, en primer lugar, razones que deriven de una simpatía innata y de un interés por los demás que la evolución ha incorporado en ti. Además, las consideraciones de autointerés pueden llevarte a desarrollar preocupaciones no autointeresadas. Por una parte, hay una cuestión de eficiencia. Te es útil tener la buena reputación de que eres alguien en quien se puede confiar. Si la gente no pensara que vas a devolver lo que te han prestado, no te daría dinero; si no pensara que eres honesto, no confiaría en ti; si no pensara que eres amable y generoso, no se sentiría amable y generosamente dispuesta hacia ti. Pero una manera de tener una buena reputación es volverse el tipo de persona cuya reputación quieres tener. Si conviertes en una regla ser siempre honesto cuando tratas con los demás, pagar siempre tus deudas, ser amable y generoso, ayudar a los demás cuando precisan ayuda, etc., es probable que los otros lo adviertan y es probable que adquieras la reputación que quieres. Ésta puede ser muy bien la mejor manera de adquirir ese tipo de reputación; dado que es difícil estar calculando todo el tiempo, y dado que cualquier desliz puede deshacer completamente meses e incluso años de planeación cuidadosa. En consecuencia, tienes una razón poderosa para actuar moralmente aun en los casos en los que parece poco probable que alguien pudiera llegar a saber que no estabas actuando así. Pero esto equivale a decir que tienes una razón autointeresada para desarrollar un interés para actuar moralmente, aun cuando actuar moralmente no favorezca tu autointerés en una ocasión particular.

También hay razones autointeresadas de un tipo diferente para que uno desarrolle intereses intrínsecos por los demás. La vida de alguien

que se preocupa por los demás es probablemente una vida más feliz que la vida de alguien que es primordialmente autointeresado. Si te preocupas por los demás, entonces tu vida será más interesante, más variada, más excitante. Su felicidad conducirá a tu felicidad. No estarás sujeto a la agonía destructiva de la competencia. Al competir con los demás, si tú y ellos son autointeresados, alguien gana y los demás pierden. En consecuencia, la mayoría es infeliz. Si, en cambio, tú y los demás no se preocupan únicamente por sí mismos, sino también por el otro, entonces no hay una verdadera competencia. Si alguno gana, todos ganan, porque todos estarán felices con el éxito de esa persona.

6. *¿Cómo pueden surgir deseos no autointeresados?*

Lo que le molesta al egoísta acerca de esto, aunque crea en ello, es que no entiende cómo alguien podría desarrollar genuinamente preocupaciones no autointeresadas. El egoísta arguye, por ejemplo, que si desarrollas una preocupación por los demás porque hacerlo redunda en tu interés, tu preocupación tiene que ser una preocupación por tu propio interés y no puede ser una preocupación genuinamente no autointeresada por alguien más.

Pero esto es suponer erróneamente que, si configuras un deseo de F porque quieres E, tu deseo de F es a fin de cuentas un deseo de E; esto es suponer que puedes formarte tal deseo de F sólo porque crees que F te conducirá a E. Sin embargo, hay otros casos que el egoísta descuida. Puedes, al menos por un lapso breve, interesarte por algo o conformar un deseo de algo, no por lo que esperas obtener de ello, sino por lo que esperas obtener de tu interés en ello. Puedes querer F porque quieres E, en tanto esperas obtener E no a partir de la obtención de F, sino de querer obtener F. Es fácil pasar esto por alto porque es un fenómeno común. Se manifiesta todo el tiempo en las competencias deportivas y en otras formas de juego. Temporalmente adoptas el deseo de ganar, porque es más divertido jugar si tratas de ganar que jugar cuando no tratas de hacerlo. Te interesas en un tema de conversación, no porque la información concerniente a esa materia vaya a mejorar tus intereses previos, sino porque obtienes más placer de una conversación si estás interesado en lo que se discute.

Al margen de los motivos puramente autointeresados, podrías desarrollar una preocupación a largo plazo por los demás. No es difícil

interesarse en alguien por sí mismo; lo hacemos todo el tiempo, aunque sólo sea para tornar más soportables nuestros contactos con los demás. El problema consiste en mantener tal interés a lo largo del tiempo. Tienes que repetir los ademanes una y otra vez hasta que se vuelvan habituales. Entonces, a través del mismo principio psicológico de generalización que lleva al avaro a volverse habitualmente interesado en el dinero por sí mismo en vez de interesarse en lo que puede comprar con el dinero, puedes encontrarte habitualmente interesado en los demás, por sí mismos, y no por tu propio beneficio. Paradójicamente, al abandonar tu egoísmo y al volverte interesado en los demás, actúas en tu propio interés y ayudas a hacerte feliz.

Un niño en proceso de crecimiento adquiere con unos pocos cálculos conscientes muchas preocupaciones no autointeresadas. Esto es natural. Es natural para los niños jugar, ser amados por sus padres y desarrollar otros intereses. El egoísta acierta cuando considera que el desarrollo de esos intereses redunda en el interés del niño; puede incluso estar en lo correcto al pensar que el niño desarrolla esos intereses porque le conviene hacerlo; pero éste no es un argumento a favor del egoísmo. No se sigue que esos intereses sean, en última instancia, preocupaciones por los intereses propios del niño. El niño —quizá por razones autointeresadas, quizá por otras razones— desarrolla preocupaciones genuinamente no autointeresadas. Habiendo desarrollado estas preocupaciones, el niño y el adulto en el que se transforma pueden tener razones genuinamente no autointeresadas para hacer cosas, razones que pueden incluso convertir en racional, dadas las preocupaciones presentes de la persona, hacer cosas que no redundan en su interés. El egoísmo no es la única motivación humana. No todas las razones son razones de autointerés.

7. *¿Pueden ser autointeresadas las razones morales?*

Esto no muestra por sí mismo que las razones morales sean no autointeresadas. La gente puede tener razones de autointerés para adherirse a ciertas convenciones. Si éstas son convenciones morales, las cuales dan a los participantes razones para hacer cosas, podría argüirse que ciertos participantes podrían tener razones morales para hacer cosas que en última instancia son, para ellos, razones de autointerés. Otros participantes podrían adherirse a las mismas convenciones por amor y por su preocupación por los demás; para ellos, las razones morales no serían razones de autointerés. Por otra parte, podría no ser

posible adherirse genuinamente a ciertas convenciones sin adquirir preocupaciones intrínsecas por los demás. La convención de respetar a los participantes en una convención podría ser un ejemplo. Una moralidad que involucra tal respeto ofrecería necesariamente razones morales que no serían de autointerés, incluso en el caso de la gente que participase teniendo razones de interés propio para hacerlo.

Es, además, una hipótesis verosímil acerca de nuestro uso de la palabra "moral" que las convenciones sean denominadas correctamente convenciones morales sólo si son convenciones de respetar a los participantes. Esto explicaría por qué las convenciones del lenguaje y las reglas de tránsito no se consideran propiamente parte de la moralidad. Dadas esta hipótesis y la teoría de la moralidad basada en la convención social, se seguiría casi por definición que las razones morales están basadas en preocupaciones por los demás y no en el autointerés.

Esta conclusión es intuitivamente plausible por sí misma, al margen de consideraciones teóricas. Ordinariamente suponemos que alguien que actúa únicamente por autointerés no actúa por una razón moral. El ejemplo de Kant es el de un comerciante que da el vuelto correcto porque favorece su interés que se sepa que tiene esa costumbre. El comerciante hace lo correcto, pero normalmente no pensaríamos que hizo tal cosa por razones de moralidad. Actuar por razones de moralidad es actuar no por autointerés, sino por una preocupación intrínseca y por respeto a las personas en tanto "fines en sí mismas".

NOTA ACERCA DE LECTURAS ADICIONALES

Una crítica clásica al hedonismo egoísta se encuentra en *Fifteen Sermons Preached at the Rolls Chapel* del obispo Joseph Butler, especialmente en el prefacio de Butler y en los sermones I y II.

También hay un análisis útil del hedonismo egoísta en Richard B. Brandt, *Ethical Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1959, pp. 307–314.

C.G. Hempel discute "Empiricist Criteria of Cognitive Significance" en *Aspects of Scientific Explanation*, Free Press, Nueva York, 1965.

La teoría posterior de Sigmund Freud sobre la motivación está desarrollada en "Más allá del principio del placer", *Psicología de las masas*, Alianza Editorial, Madrid, 1969.

Michael Scriven describe la manera en que el autointerés puede conducirnos a la adopción de actitudes morales en *Primary Philosophy*, Mc Graw-Hill, Nueva York, 1956, cap. 7.

XIII

EL UTILITARISMO

1. *Las razones utilitaristas*

Las razones morales no son razones de autointerés. Derivan de una preocupación intrínseca o de respeto por los demás, así como por uno mismo. He señalado que son ciertas convenciones sociales las que requieren este interés o respeto. Por otra parte, Kant sostuvo que esto es exigido por la razón práctica pura. Kant pensaba que un ser racional tiene que sentir tal interés o respeto por otros seres racionales, en virtud de ser racional. No obstante, Kant nunca fue capaz de probar que esto es así y se nos puede disculpar por dudar de ello. Sin embargo, otra teoría podría plantear que la preocupación o el respeto por los demás deriva de una simpatía innata que tiene que ser explicada en sí misma en términos de evolución.

¿Qué tipos de cosas podrían la preocupación o el respeto por los demás darnos razones para hacer? La teoría más simple y famosa es el *utilitarismo*, de acuerdo con la cual tienes una razón moral para hacer todo lo que mejor promueve el bienestar general. El utilitarismo suele describirse como la teoría según la cual siempre deberías actuar a modo de maximizar la utilidad social, donde "utilidad social" es simplemente otro nombre del bienestar general. Esta manera de expresar las cosas sirve provisoriamente aunque, tal como veremos, posiblemente sea engañosa. Sin embargo, al margen de las argucias verbales, la idea básica que subyace en el utilitarismo quedará en claro.

Se dice a menudo que el utilitarismo choca con nuestros puntos de vista ordinarios. Sin embargo, este conflicto es menor de lo que puede parecer al comienzo. Por lo demás, en los casos en que hay conflicto, el utilitarismo puede parecer, después de pensarlo bien, más razonable que nuestras convicciones usuales.

Consideremos, por ejemplo, las promesas. Nuestro punto de vista ordinario dice que prometer es ponernos a nosotros mismos bajo la obligación de hacer lo que hemos prometido hacer. La obligación no es absoluta; hay casos en los que uno está justificado en hacer una promesa mentirosa, y hay casos en los que, habiendo hecho una promesa sincera, surgen circunstancias en las que se debería hacer otra cosa que cumplir la promesa. No obstante, nuestro punto de vista ordinario dice que las promesas producen una diferencia moral. El hecho de que hayamos prometido hacer algo es, por lo menos, una razón para hacerlo, aunque esa razón sea ocasionalmente superada [*overridden*] por otras razones. Sin embargo, puede parecer que el utilitarismo no es compatible con esto. El utilitarismo es, según parece, un punto de vista que "mira hacia adelante" más que "mirar hacia atrás". Lo que es importante, según el utilitarismo, no es lo ocurrido en el pasado, sino lo que podría ocurrir en el futuro en tanto consecuencia de distintas cosas que uno podría hacer. Parecería ser irrelevante lo que una vez prometiste hacer; ahora tendrías que hacer lo que en este momento promueve mejor la utilidad social, aunque hayas prometido hacer algo muy diferente.

Las promesas pueden producir, sin embargo, una diferencia, aun para un utilitarista, debido a que las promesas en sí mismas tienen consecuencias. Ocurre que existe en nuestra sociedad una práctica que consiste en cumplir las promesas que uno ha hecho. Cuando prometes hacer algo, en consecuencia, tiendes a crear, en la persona a quien le has hecho la promesa, una expectativa de que harás la cosa en cuestión, y ella ajustará sus planes de acuerdo con ello. En ese caso, si no haces lo que has prometido hacer, la desilusionarás, y frustrarás sus planes. Si no hubieras hecho tu promesa, el otro no estaría contando contigo. Por eso, las consecuencias de que dejes de hacer algo, a menudo serán diferentes si has prometido, o no, hacerlo; de tal modo, sería incorrecto decir que, de acuerdo con el utilitarismo, las promesas no producen ninguna diferencia moral. Sí la producen porque generan una diferencia en las consecuencias de las acciones posteriores y, según el utilitarismo, las consecuencias producen una diferencia moral.

Seguir adhiriéndose a la práctica de cumplir las promesas es también útil por otros motivos: anima a la gente a confiar en uno en el futuro. Si los demás dejaran de confiar en ti con respecto al cumplimiento de tus promesas, se te haría difícil coordinar tus acciones con las de ellos, y las cosas que podrías hacer para promover la uti-

lidad se verían muy reducidas. Al considerar si deberías cumplir una promesa particular, por supuesto, tienes que considerar no sólo los efectos directos de los distintos cursos de acción que puedas iniciar, sino además todos los efectos indirectos, por ejemplo, en que en el futuro puedas continuar coordinando tus acciones con las de los demás. En la mayoría de los casos, esta consideración es suficientemente importante como para darte una razón utilitarista poderosa para cumplir tu promesa.

2. *Explicación de las intuiciones no utilitaristas*

Los dos factores que he mencionado tienen que ver con las expectativas de la persona a la que has hecho tu promesa y, además, con tu capacidad de continuar usando las promesas para coordinar en el futuro tus acciones con las de los demás; por ello, no están involucrados en los casos que a veces se denominan "promesas en la isla desierta". Supongamos, por ejemplo, que tú y otra persona han naufragado en una isla desierta y que la otra persona está muriendo. Antes de morir te da una suma de dinero y tú prometes que, si eres rescatado, darás el dinero a cierto individuo, Austin Jones, quien es su hijo ilegítimo no reconocido. Tú prometes; él muere contento; y poco después eres rescatado. Al volver a tu país, descubres que Jones es un hombre rico pero extremadamente egoísta y comprendes, inmediatamente, que sería mucho mejor si el dinero fuera entregado a alguna obra de caridad útil antes que a Jones. ¿Qué harías?

Ciertamente no vas a frustrar las expectativas de nadie. La única persona que supo de tu promesa está ahora muerta. En consecuencia, lo que harás no afectará su bienestar. Jones, por otra parte, no sabe nada acerca de tu promesa y en consecuencia no tiene ninguna expectativa que pueda ser frustrada si no le das el dinero. Además, dado que nadie sabe que has hecho esa promesa, no cumplirla no afectará la confianza futura en el cumplimiento de tus promesas. Podrías incluso mentir y decir que has prometido que cuidarías que el dinero se destinara a la caridad. Al dar el dinero a la caridad, en ese caso, reforzarías la confianza que los otros tienen en tu confiabilidad. En consecuencia, parecería que, desde un punto de vista utilitarista, que hayas hecho una promesa sería totalmente irrelevante. Deberías hacer lo que es mejor y esto es lo que habría sido mejor si no hubieras hecho tu promesa. Deberías dar el dinero a Jones, tal como lo prometiste, sólo si también ésa hubiera sido la acción correcta en cuanto al dinero, aunque no lo

hubieras prometido. Y como darle el dinero a Jones no habría sido lo correcto, tendrías que romper tu promesa y destinar el dinero a la caridad.

Aquí, sin embargo, hay un conflicto claro entre el utilitarismo y la opinión moral ordinaria. Corrientemente supondríamos que una promesa, incluso una promesa en una isla desierta, hecha a un hombre moribundo, y acerca de la cual nadie sabe, tiene algún peso moral. Sin duda, suponemos ordinariamente que la promesa no tiene un peso absoluto, sino *algún* peso. Pero, según el utilitarismo, en ese caso la promesa no tiene peso. Esto es antiintuitivo.

La contestación utilitarista a esta objeción consiste en estar de acuerdo en que la conclusión utilitarista es antiintuitiva, y en sostener que, no obstante, es correcta. La conclusión utilitarista es antiintuitiva en el sentido de que toda persona común sentirá en este caso alguna motivación para hacer lo que ha prometido hacer, a saber, dar el dinero a Jones, y puede incluso sentirse culpable si en lugar de ello, destina el dinero a la caridad. Esto ocurre porque ha sido entrenado para cumplir sus promesas y sentirse culpable si no lo hace. Además, hay una buena razón utilitaria para entrenar a la gente de esta manera en vez de tratar de entrenarla para actuar como utilitarista racional. Pues, en muchos casos, no está claro qué es correcto hacer desde un punto de vista utilitario. En consecuencia, es útil tener una regla empírica para decidir qué hacer en vez de que cada persona tenga que comenzar de nuevo en cada caso. Si además la gente fuera entrenada para decidir qué hacer en cada caso, a medida que se produce, probablemente cometería muchos errores desastrosos. Las consideraciones utilitarias importantes son a menudo complejas, indirectas y fácilmente pasadas por alto; por ejemplo, los efectos indirectos del cumplimiento de las promesas ya advertidos. En consecuencia, es útil entrenar a la gente para que siga reglas generales, salvo cuando resulta absolutamente claro que la utilidad sería maximizada si se actuara de otra manera. Una vez que la gente esté entrenada así para seguir reglas generales, puede tener intuiciones que ocasionalmente entren en conflicto con el utilitarismo y puede sentirse culpable por actuar en contra de ellas incluso si, como buen utilitarista, sabe que está haciendo lo que realmente es correcto hacer. Por lo tanto, aun cuando el utilitarismo está en conflicto con los puntos de vista ordinarios acerca de las promesas en la isla desierta, puede dar una explicación de nuestros sentimientos corrientes hacia tales casos, una explicación de por qué podríamos

racionalmente tener esos sentimientos aunque la acción correcta sea diferente de lo que ordinariamente consideramos que sería.

Otras intuiciones no utilitaristas podrían explicarse de manera similar. Reconocemos, por ejemplo, una distinción moral intuitiva entre dañar a alguien y no ayudar a alguien; pensamos que un médico no despedazaría a uno de sus pacientes para salvar a otros cinco pacientes, aunque, desde un punto de vista utilitarista, eso es precisamente lo que el médico debería hacer. El utilitarista tendrá que decir que nuestro juicio natural intuitivo acerca de este ejemplo es erróneo; puede, no obstante, explicar por qué tenemos tales intuiciones y por qué es racional, desde un punto de vista utilitarista, entrenar a la gente de modo que tenga tales reacciones no utilitaristas. Ya que, en muchos casos, distinguir dañar de no ayudar tiene un sentido utilitarista, el cual radica en que lo que ayudará a alguien depende frecuentemente de las metas particulares de esa persona, y las metas varían de una persona a otra. Merodear tratando de ayudar a la gente causa con frecuencia más mal que bien; a menudo, es interferir en vez de ayudar. Por otra parte, suele resultar bastante claro que un curso dado de acción causará daño a alguien. En consecuencia, como una buena regla empírica, hay a menudo una razón utilitarista más fuerte para tratar de evitar dañar a los demás que para tratar de ayudarlos. Ayudar a los demás es bueno desde un punto de vista utilitarista, pero lo que constituye ayuda no siempre es tan claro como lo que constituye daño. Quien trata de ayudar a los demás corre el riesgo de ser un entrometido que interfiere, o algo aún peor. Al no conocer las metas y los objetivos de otra persona, uno puede incorrectamente proyectar sobre esa persona sus propias metas y objetivos. Las consideraciones de autointerés por parte de uno pueden producir una distorsión adicional. En consecuencia, desde un punto de vista utilitarista, tú y los demás tendrían que ser entrenados a manera de que estuvieran fuertemente motivados a evitar las acciones que podrían dañar a los demás y a tratar de ayudarlos sólo cuando resulta muy claro que se está ayudando y no sólo interfiriendo. Dado este entrenamiento, sin embargo, se formulará una distinción intuitiva inmediata entre dañar a alguien y no ayudarlo. Por ejemplo, se juzgará que el médico no tendría que dañar a un paciente para salvar a los otros cinco. Este juicio es incorrecto desde un punto de vista utilitarista. No obstante, hay una explicación utilitarista de por qué habrías de ser entrenado para hacerlo.

3. *¿Requiere demasiado el utilitarismo?*

Hemos estado considerando la manera en que el utilitarismo podría responder la objeción de que algunas veces recomienda cursos de acción que son intuitivamente incorrectos. También se formula una objeción opuesta: el utilitarismo nos pide demasiado. Porque el utilitarismo dice que siempre deberías llevar a cabo el acto que maximice la utilidad, y es extremadamente raro que cualquiera haga eso alguna vez. Consideremos tu situación actual. Estás leyendo un libro filosófico acerca de la ética. Hay cursos de acción abiertos a ti que podrían tener una utilidad social mayor. Si, por ejemplo, dejaras inmediatamente de leer e hicieras algo que pudiera permitirte enviar comida a lugares tales como África o India, donde es escasa, podrías probablemente salvar cientos o incluso miles de vidas y podrías hacer la vida más tolerable para miles de personas. Esto es algo que podrías hacer que tiene una utilidad mayor que lo que estás haciendo ahora; probablemente tiene una utilidad mayor que lo que has hecho en toda tu vida. En consecuencia, de acuerdo con el utilitarismo, no estás haciendo ahora lo que moralmente deberías estar haciendo y esto seguirá siendo verdad a lo largo de toda tu vida. Tú siempre estarás haciendo lo incorrecto; nunca estarás haciendo lo que deberías estar haciendo. Esa conclusión es muy diferente de cualquier visión ordinaria de la cuestión. El utilitarismo implica que es moralmente incorrecto de tu parte estar leyendo este libro. Éste no es el tipo de cosas que la gente piensa ordinariamente como moralmente incorrectas.

Alguien que actuara de la manera recomendada por el utilitarismo no sería un buen ejemplo de una persona que hiciera lo que uno moralmente debería hacer, tal como ordinariamente concebimos a dicha persona. Un agente utilitarista sería muy diferente de la mayoría de nosotros. Sería un santo o un super Ralph Nader. Dicha persona haría en todo momento lo que le parecería tener la mejor posibilidad de maximizar la utilidad. Ocasionalmente descansaría; pero sólo con objeto de estar en condiciones de volver a la batalla con renovado vigor. Este cuadro inspira, pero no es la imagen que tenemos de quien hace lo que debería hacer; es nuestra imagen de alguien que hace mucho más.

El teórico utilitarista podría contestar que el hecho de que seamos débiles y holgazanes y de que no deseemos hacer lo que es claramente deseable no es un argumento contra el utilitarismo. No es un argumento contra la suposición de que moralmente deberíamos hacer lo que es claramente deseable hacer. Sería claramente mejor estar trabajando

activamente para reducir y eliminar el hambre, en vez de estar leyendo filosofía. Moralmente deberías estar trabajando para reducir el hambre y es moralmente incorrecto de tu parte continuar leyendo este libro, aunque eres demasiado holgazán para hacer lo que moralmente deberías hacer.

Sin embargo, esta respuesta utilitarista no es totalmente satisfactoria. Podemos concordar en que lo que el utilitarismo recomienda es o sería espléndido. Sería espléndido si todos nosotros actuáramos de la manera en que el utilitarista dice que deberíamos actuar. Pero ¿significa esto que es un *requisito* moral que actuemos de esa manera? El punto es que el utilitarismo parece anular ciertas distinciones. Ordinariamente suponemos que hay diferencia entre decir que sería espléndido por parte de *P* hacer *H*, que sería agradable por parte de *P* hacer *H*, que *P* debería hacer *H*, que *P* está obligado a hacer *H*, que *P* tiene el deber de hacer *H* y que *P* tiene que hacer *H*. Ordinariamente suponemos, por ejemplo, que hay circunstancias en las cuales sería agradable que alguien hiciera algo aunque no sea el caso que tenga el deber de hacerlo o incluso que debería hacerlo. Sería muy agradable que me pagaras el almuerzo; sin embargo, no estás obligado a hacerlo y ni siquiera es verdadero que deberías hacerlo. Por otro lado, si tú y yo regularmente pagamos el almuerzo del otro y hoy yo no tengo dinero, quizá deberías pagar mi almuerzo aunque no tengas la obligación o el deber de hacerlo. Aquí hay distinciones del tipo que hacemos todo el tiempo en nuestro pensamiento moral ordinario; parecería que no hay cabida para ellas en el pensamiento utilitario, si la única cuestión es maximizar la utilidad. Puede parecer que, para un utilitarista, todo lo que maximiza la utilidad tiene que ser simultáneamente lo que sería espléndido hacer, lo que sería agradable hacer, lo que uno debería hacer, lo que uno tiene que hacer, lo que uno está obligado a hacer, lo que uno tiene el deber de hacer. No es de extrañar que el utilitarismo haya parecido a muchos filósofos una teoría extremadamente tosca.

Ya he mencionado otra distinción que ordinariamente realizamos y que el utilitarista cree que tiene alguna significación sólo como una regla empírica práctica: la distinción moral entre dañar a alguien y no ayudarlo. En este caso, nuevamente, la manera utilitarista de hablar parece llevarnos a conclusiones curiosas. Según la moralidad utilitarista, que no seamos capaces de salvar la vida de alguien parece no ser diferente de matar efectivamente a alguien. Desde un punto de vista utilitarista, dado que cualquiera de nosotros podría, si trata-

ra, salvar la vida de muchas personas, al no hacer lo que podemos para salvar esas vidas aparentemente no somos mejores que los asesinos. El utilitarista puede suponer que, desde el punto de vista moral, no somos diferentes de los jefes de un campo de concentración que ponen en las cámaras de gas a cientos de prisioneros.

En consecuencia, puede parecer que la moralidad utilitarista resulta ser una disciplina curiosamente abstracta con muy poco de la complejidad de nuestro pensamiento moral ordinario: todos somos asesinos de masas de prisioneros; siempre actuamos de manera incorrecta; nunca hacemos lo que moralmente deberíamos hacer.

4. *Una explicación utilitarista de nuestras distinciones morales ordinarias*

Pero, de hecho, la explicación utilitarista de la moralidad puede permitirnos formular la mayoría de las distinciones morales que ordinariamente querríamos formular, al considerar cuándo, desde un punto de vista utilitarista, sería apropiado premiar o censurar a alguien. Ordinariamente distinguimos, por ejemplo, entre las cosas que una persona debería hacer y otras cosas que no es el caso que debería hacer pero que sería agradable de su parte hacer o espléndido si las hiciera. El utilitarista podría explicar esto como una distinción, dentro de la clase de cosas que una persona debería hacer, entre las cosas que merecen premio si se hacen y cosas que no merecen premio si se las hace. De manera similar, para algunas cosas pero no para otras que tú deberías hacer, deberías ser censurado si no las hicieras. En el caso en el que ordinariamente diríamos que *P* debería hacer *H*, el utilitarista podría simplemente decir que *P* merecerá censura por no hacer *H*. En el caso en que dijéramos que sería agradable por parte de *P* hacer *H*, aun cuando *P* no tenga que hacerlo, el utilitarista podría decir simplemente que *P* merecerá un premio moderado si *P* hace *H*. En el caso en que vislumbramos una distinción moral entre alguien que lee este libro, dejando de salvar con ello cientos de vidas al no trabajar para las instituciones de bien público, y la persona que dirige un campo de concentración y mata a cientos de prisioneros, el utilitarista puede ver una diferencia en el tipo de condenación que merecen estos actos distintos. En cualquier caso, el utilitarista podría decir que la gente está haciendo algo que no debería estar haciendo, pero él bien puede identificar una razón utilitarista para querer condenar con más fuerza las acciones de la persona que dirige el campo de concentración, que

las de la persona que está leyendo este libro en lugar de trabajar para socorrer a los hambrientos.

El utilitarismo concibe la moralidad como una cuestión de benevolencia. La moral intuitiva es altruista y se preocupa por los demás. En un estado de cosas ideal, todos serían perfectamente benevolentes; todos considerarían los intereses de los demás tan importantes como los propios. Sin embargo, la gente no es así; y es posible que nunca lo sea. El propósito de la alabanza y la censura, desde un punto de vista utilitario, es animar a la gente a que actúe de una manera más benevolente y menos autointeresadamente de como lo hace. A la gente le gusta ser alabada y no le gusta ser censurada. En consecuencia, está motivada para actuar de manera que sea alabada y no censurada. Pero ésta es sólo una motivación. Hay otras. La gente está motivada a actuar de modo que sea alabada y no sea censurada, si puede hacerlo con facilidad y sin mayores problemas. Si premiamos a la gente sólo cuando ha actuado a fin de maximizar la utilidad esperada y censuramos a la gente siempre que no lo hace, nuestra alabanza y nuestra censura serían antiproductivas. Pues recibir alabanzas y evitar la censura requeriría demasiado esfuerzo por parte de la gente, se perdería interés en tratar de hacerlo. Tal como la gente es en la actualidad, ha alcanzado un pacto entre las pretensiones de ser benevolente y más autointeresada. La gente es autointeresada y, en alguna medida, benevolente. El propósito de la alabanza y la censura es alentar a la gente a volverse más benevolente y menos autointeresada. Hacemos esto alabando las acciones que son más benevolentes de lo que usualmente esperamos de la gente y censurando las acciones que son menos benevolentes y más autointeresadas que aquellas que usualmente esperamos de la gente. De esta manera, esperamos coaccionar a la gente para que se dirija hacia los fines morales de las cosas y se aparte de los fines autointeresados. Pero todo esto depende de que no ubiquemos los estándares de alabanza y de censura demasiado alto. De otra manera, la gente no haría caso de nuestra alabanza y nuestra censura, dado que sería demasiado difícil para ella ser premiada y no ser censurada.

5. *El utilitarismo como ideal*

He estado escribiendo como si el utilitarismo fuera expresado mejor como la teoría según la cual uno moralmente debería hacer lo que maximizara la utilidad esperada. Pero ahora aparece una terminolo-

gía alternativa. Podríamos decir que lo que alguien debería hacer, de acuerdo con el utilitarismo, es todo aquello para lo que haya razones utilitaristas para censurarlo por no hacerlo; sería agradable (o bueno) de parte de alguien hacer algo para lo que hay razones utilitaristas para alabarlo; etc. Esta terminología permite al utilitarista distinguir las cosas que sería agradable que alguien hiciera, de las cosas que debería hacer, usando esas palabras de manera similar a la de su sentido ordinario.

Lo que el utilitarismo dice que nosotros deberíamos hacer puede, en consecuencia, estar más próximo a nuestras opiniones usuales de lo que parecía al principio. Porque, cuando se reformula el utilitarismo de la manera que he sugerido, no está comprometido de manera obvia a decir que el médico debería despedazar a un paciente para salvar a los otros cinco, porque censurar al médico por no hacerlo sería casi con certeza antiintuitivo. Tampoco está comprometido a decir que haces algo incorrecto cuando lees este libro en vez de estar trabajando por la eliminación del hambre. Estás actuando como lo hace la mayoría de la gente; por lo tanto, probablemente se gana muy poco censurándote por lo que haces. Probablemente sería mejor alabar a quienes tratan de eliminar el hambre que censurar a quienes no lo hacen.

Pero, aunque el utilitarismo no se opone obviamente a los puntos de vista que consideramos morales en lo que dicen que deberíamos hacer, hay un aspecto importante en el que no ofrece una explicación adecuada de nuestra moralidad. El utilitarismo podría interpretarse como la teoría según la cual las razones morales se derivan enteramente de una preocupación por el bienestar general. Queda claro, no obstante, por la manera en que tiene que defenderse el utilitarismo, que podemos tener razones morales que deriven de otra preocupación. Al explicar nuestros sentimientos intuitivos acerca de las promesas en una isla desierta, por ejemplo, el utilitarismo advierte que cada uno de nosotros ha sido educado para preocuparse, intrínsecamente, por el cumplimiento de sus promesas. Este interés intrínseco puede ofrecer a alguien una razón para cumplir una promesa al margen de las consideraciones referidas al bienestar general. Tal razón es, con seguridad, una razón moral. En consecuencia, es una razón moral que no surge de una preocupación por el bienestar general.

Nuestra adhesión al cumplimiento de las promesas no es irracional, desde un punto de vista utilitarista, porque promueve el bienestar general. El utilitarista no puede decir, en consecuencia, que deberíamos

abandonar esa preocupación que nos da razones morales que no se derivan de una preocupación por el bienestar general. Además, el hecho de que exista esa justificación utilitarista de nuestra preocupación no significa que la preocupación misma sea efectivamente una preocupación por el bienestar general, como tampoco el hecho de que haya una justificación de esta preocupación en términos de autointerés significa que la preocupación misma sea una preocupación autointeresada.

Aceptamos muchas otras convenciones morales de la manera en que aceptamos las convenciones asociadas a las promesas. Tenemos una preocupación intrínseca inmediata para actuar de acuerdo con esas convenciones. El estímulo originario para que desarrolláramos esa preocupación probablemente fue el autointerés. Nos interesaba desarrollar esa preocupación y que se supiera que lo habíamos hecho. Dado que el bienestar general es simplemente la suma de los intereses de todos, también hay una razón utilitarista para que sigamos teniendo esa preocupación. Pero, nuevamente, esto no significa que nuestra preocupación moral sea autointeresada o incluso, típicamente, una preocupación por el bienestar general.

Las razones morales se derivan no sólo de nuestra preocupación por los demás sino también de nuestro respeto a ellos. ¿Qué significa eso? Kant sugirió que esto significa respeto por un derecho en el que tanto ellos como nosotros hemos concordado. Podríamos decir que es un respeto por las convenciones morales de nuestra sociedad. Esas convenciones definen nuestra concepción del respeto. No comemos carne humana; hacerlo en nuestra sociedad indicaría que no respetamos a la persona que hemos comido y a la gente en general. Para nosotros, dañar a alguien intencionalmente es mostrar una falta de respeto por él, que no muestra nuestra falta de ayuda, en igualdad de circunstancias. De hecho, los intentos por ayudar a la gente en nuestra sociedad a menudo se interpretarán como una falta de respeto por la gente involucrada. Por ejemplo, mucha gente se siente "demasiado orgullosa" de estar a favor del bienestar. Las cosas serían diferentes en una sociedad más igualitaria que no distinguiera moralmente dañar de dejar de ayudar a alguien, tal como lo hacemos. En tal sociedad, no ser capaces de ayudar a alguien sería un signo de falta de respeto, tal como lo es dañar a alguien.

El respeto por los demás involucra alguna preocupación por ellos. Por lo tanto, hay una presión a favor del utilitarismo. El autointerés nos lleva a adoptar las convenciones de respeto y de preocupación. Nuestra preocupación por los demás nos dará entonces razones para

mejorar nuestras convenciones a fin de promover mejor el bienestar general. De este modo progresamos moralmente; nuestro autointerés da lugar a la benevolencia de la manera en que aconseja el utilitarismo. Pese a ello, permanecemos autointeresados. Eso sujeta las riendas de la moralidad y nos impide recorrer el camino en la dirección de la benevolencia y del altruismo unilaterales.

NOTA ACERCA DE LECTURAS ADICIONALES

Los enunciados clásicos del utilitarismo pueden encontrarse en *Introduction to the Principles of Morals and Legislation* de Jeremy Bentham y *Utilitarianism* de John Stuart Mill.

Para una defensa más reciente, véase J.J.C. Smart, "Extreme and Restricted Utilitarianism", *Philosophical Quarterly*, vol. 6, 1956; y sus contribuciones a J.J.C. Smart y Bernard Williams, *Utilitarianism: For and Against*, Cambridge University Press, Cambridge, 1973.

ÍNDICE DE AUTORES

Anscombe, G.E.M 78
Aristóteles 43, 81–2, 84–5, 144
Ayer, A.J. 36

Baier, Kurt 140
Benacerraf, Paul 21
Bentham, Jeremy 180
Brandt, Richard B. 67, 167
Butler, Joseph 167

Davidson, Donald 140
Dostoievsky, Fyodor 23, 35

Firth, Roderick 67
Foot, Philippa 36, 53, 78, 140
Frankena, William K. 36, 115
Freud, Sigmund 74, 78, 156, 159–60, 167

Gewirth, Alan 21

Hanson, Norwood Russell 21
Hare, R.M. 20, 53, 94–105, 108, 128, 131–3, 136, 140, 143, 149
Harman, Gilbert 21, 130
Hart, H.L.A. 78
Hempel, C.G. 167
Hume, David 41–2, 50–2, 53, 58, 64, 81–2, 84–5, 119–21, 125–6, 130, 154, 164

Isenberg, Arnold 53

Kant, Immanuel 79–91, 93–4, 121, 125–6, 128, 131, 137, 148–9, 167, 169, 179
Katz, Jerrold J. 36

Lewis, David 130

Mill, John Stuart 180
Moore, G.E. 36

Nagel, Thomas 84–6, 91
Nietzsche, Friedrich 35

Piaget, Jean 67

Ross, W.D. 135–6, 140

Sartre, Jean-Paul 94–5, 98, 102, 105, 131
Scriven, Michael 167
Skinner, B.F. 161
Smart, J.J.C. 180
Stevenson, Charles L. 52
Strawson, P.F. 21

ÍNDICE GENERAL

PREFACIO 7

1. EL PROBLEMA CON LA ÉTICA

I. ÉTICA Y OBSERVACIÓN

1. La dificultad básica 13
2. La observación 14
3. Pruebas basadas en la observación 16
4. La ética y la matemática 20

II. NIHILISMO Y NATURALISMO

1. El nihilismo moral 23
2. Reducciones 25
3. El naturalismo ético: el funcionalismo 26
4. El argumento de la pregunta pendiente [*open question*] .. 30
5. El naturalismo redefinicional 32
6. Por qué es problemática la ética 33

2. EL EMOTIVISMO

III. EL EMOTIVISMO COMO NIHILISMO MODERADO

1. El emotivismo: la idea básica 39
2. El emotivismo y el argumento de la pregunta pendiente .. 41
3. Ventajas del emotivismo 45
4. La verdad en la ética 46
5. El razonamiento moral 48
6. La escapatoria de Hume 50
7. La trivialidad del emotivismo 52

IV. EL EMOTIVISMO COMO TEORÍA DEL OBSERVADOR IDEAL

1. La teoría del observador ideal 55
2. El relativismo moral y la teoría del observador ideal 58
3. El emotivismo *versus* la teoría del observador ideal 60
4. La amenaza de circularidad 62
5. Los principios morales y el observador ideal 63
6. Una moraleja 66

3. LA LEY MORAL

V. SOCIEDAD Y SUPERYÓ

1. Las reglas morales 71
2. La sanción moral 73
3. ¿Puede evaluarse la moralidad? 76

VI. LEYES DE RAZÓN

1. La fuente de la motivación moral 79
2. ¿Por qué preocuparse por otro? 82
3. Los deseos en tanto datos 84
4. La prueba de Kant 86
5. Objeciones a Kant 90

VII. PRINCIPIOS PERSONALES

1. Una ley moral personal 93
2. La teoría de R.M. Hare 95
3. La objeción existencialista a Hare 98
4. Las razones y los principios básicos 102

VIII. COSTUMBRES Y RELATIVIDAD

1. El internalismo y el externalismo 107
2. La costumbre social y la moralidad 109
3. El relativismo otra vez 112

4. RAZONES Y CONVENCIONES

IX. LA CONVENCIÓN

1. Las convenciones morales 119
2. Juzgando a los extraños 123

3. Los aspectos convencionales de la moralidad 126
4. La teoría de la convención tácita y la teoría de Kant 128

X. RAZONES

1. La lógica y las razones 131
2. Un análisis en términos de buenas razones 133
3. Los "debería" *prima facie* 135
4. Algunas complicaciones 139

XI. UNA TEORÍA NATURALISTA DE LAS RAZONES

1. Las razones y el razonamiento 141
2. El razonamiento como prueba o como argumento 142
3. Una mejor explicación del razonamiento 144
4. El razonador ideal 145
5. Los hechos morales 147

5. YO Y LOS OTROS

XII. EL EGOÍSMO

1. El autointerés 153
2. El hedonismo egoísta 155
3. Una objeción estándar 157
4. Réplica a la objeción 158
5. Lo que la gente quiere 160
6. ¿Cómo pueden surgir deseos no autointeresados? 165
7. ¿Pueden ser autointeresadas las razones morales? 166

XIII. EL UTILITARISMO

1. Las razones utilitaristas 169
2. Explicación de las intuiciones no utilitaristas 171
3. ¿Requiere demasiado el utilitarismo? 174
4. Una explicación utilitarista de nuestras distinciones morales ordinarias 176
5. El utilitarismo como ideal 177

ÍNDICE DE AUTORES 181

La naturaleza de la moralidad. Una introducción a la ética, publicado por el Instituto de Investigaciones Filosóficas, se terminó de imprimir el 30 de noviembre de 2009, en los talleres de Impresión Comunicación Gráfica S.A. de C.V., situados en Manuel Ávila Camacho no. 689, colonia Santa María Aztahuacán, Delegación Iztapalapa, C.P. 09500, México, D.F. Para su impresión, realizada en offset, se utilizó papel cultural de 90 gramos; en su composición y formación, llevadas a cabo por computadora, se usaron el programa TEX y tipos Times de 10.5/12, 9.5/11 y 8.5/10. El cuidado de la edición y la formación tipográfica estuvieron a cargo de Laura E. Manríquez.
El tiraje consta de 1000 ejemplares.

El presente libro se inicia con el examen de un problema central acerca de la ética: su aparente inmunidad con respecto a las pruebas observacionales. En un estilo informal pero preciso, Gilbert Harman, profesor de filosofía en la universidad de Princeton, evalúa posibles respuestas a este problema, tales como el nihilismo extremo, el emotivismo y el naturalismo ético. A continuación aborda la manera en que la moralidad parece implicar alguna especie de ley moral y se pregunta cómo hay que entender esta última, incorporando a su discusión la teoría de Kant y después las de Hare y Sartre. Más adelante se explora la posibilidad de una fuente social de ley moral, y finalmente, tras considerar cómo puede decirse que los principios morales nos dan razones para actuar, Harman discute el egoísmo y el utilitarismo como puntos de vista sobre la naturaleza de las razones.

ISBN 978-968-36-5062-7

INSTITUTO DE INVESTIGACIONES FILOSÓFICAS

UNAM POSGRADO Filosofía